

NOVEL POLYPEPTIDES AND NUCLEIC ACIDS ENCODING SAME

5

RELATED APPLICATIONS

This application claims priority to USSN 60/188,316, filed March 10, 2000 (15966-721);
USSN 60/188,277, filed March 10, 2000 (15966-722); USSN 60/189,139, filed March 14, 2000
10 (15966-724); USSN 60/189,140, filed March 14, 2000 (15966-725); USSN 60/190,401, filed
March 17, 2000 (15966-726); and USSN 60/190,231, filed March 17, 2000 (15966-727), which
are incorporated herein by reference in their entirety.

TECHNICAL FIELD OF THE INVENTION

15 The invention generally relates to nucleic acids and polypeptides encoded therefrom.

BACKGROUND OF THE INVENTION

The invention generally relates to nucleic acids and polypeptides encoded therefrom.
More specifically, the invention relates to nucleic acids encoding cytoplasmic, nuclear,
20 membrane bound, and secreted polypeptides, as well as vectors, host cells, antibodies, and
recombinant methods for producing these nucleic acids and polypeptides.

SUMMARY OF THE INVENTION

The invention is based, in part, upon the discovery of novel polynucleotide sequences
25 encoding novel polypeptides.

Accordingly, in one aspect, the invention provides an isolated nucleic acid molecule that
includes the sequence of SEQ ID NO: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, or 21 or a fragment,
homolog, analog or derivative thereof. The nucleic acid can include, *e.g.*, a nucleic acid
sequence encoding a polypeptide at least 85% identical to a polypeptide that includes the amino
30 acid sequences of SEQ ID NO: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20 or 22. The nucleic acid can be,
e.g., a genomic DNA fragment, or a cDNA molecule.

Also included in the invention is a vector containing one or more of the nucleic acids
described herein, and a cell containing the vectors or nucleic acids described herein.

The invention is also directed to host cells transformed with a vector comprising any of the nucleic acid molecules described above.

In another aspect, the invention includes a pharmaceutical composition that includes a NOVX nucleic acid and a pharmaceutically acceptable carrier or diluent.

5 In a further aspect, the invention includes a substantially purified NOVX polypeptide, *e.g.*, any of the NOVX polypeptides encoded by a NOVX nucleic acid, and fragments, homologs, analogs, and derivatives thereof. The invention also includes a pharmaceutical composition that includes a NOVX polypeptide and a pharmaceutically acceptable carrier or diluent.

10 In still a further aspect, the invention provides an antibody that binds specifically to a NOVX polypeptide. The antibody can be, *e.g.*, a monoclonal or polyclonal antibody, and fragments, homologs, analogs, and derivatives thereof. The invention also includes a pharmaceutical composition including NOVX antibody and a pharmaceutically acceptable carrier or diluent. The invention is also directed to isolated antibodies that bind to an epitope on a polypeptide encoded by any of the nucleic acid molecules described above.

15 The invention also includes kits comprising any of the pharmaceutical compositions described above.

The invention further provides a method for producing a NOVX polypeptide by providing a cell containing a NOVX nucleic acid, *e.g.*, a vector that includes a NOVX nucleic acid, and culturing the cell under conditions sufficient to express the NOVX polypeptide encoded by the nucleic acid. The expressed NOVX polypeptide is then recovered from the cell. Preferably, the cell produces little or no endogenous NOVX polypeptide. The cell can be, *e.g.*, a prokaryotic cell or eukaryotic cell.

20 The invention is also directed to methods of identifying a NOVX polypeptide or nucleic acid in a sample by contacting the sample with a compound that specifically binds to the polypeptide or nucleic acid, and detecting complex formation, if present.

The invention further provides methods of identifying a compound that modulates the activity of a NOVX polypeptide by contacting a NOVX polypeptide with a compound and determining whether the NOVX polypeptide activity is modified.

30 The invention is also directed to compounds that modulate NOVX polypeptide activity identified by contacting a NOVX polypeptide with the compound and determining whether the compound modifies activity of the NOVX polypeptide, binds to the NOVX polypeptide, or binds to a nucleic acid molecule encoding a NOVX polypeptide.

In another aspect, the invention provides a method of determining the presence of or predisposition of a NOVX-associated disorder in a subject. The method includes providing a sample from the subject and measuring the amount of NOVX polypeptide in the subject sample. The amount of NOVX polypeptide in the subject sample is then compared to the amount of NOVX polypeptide in a control sample. An alteration in the amount of NOVX polypeptide in the subject protein sample relative to the amount of NOVX polypeptide in the control protein sample indicates the subject has a tissue proliferation-associated condition. A control sample is preferably taken from a matched individual, *i.e.*, an individual of similar age, sex, or other general condition but who is not suspected of having a tissue proliferation-associated condition. Alternatively, the control sample may be taken from the subject at a time when the subject is not suspected of having a tissue proliferation-associated disorder. In some embodiments, the NOVX is detected using a NOVX antibody.

In a further aspect, the invention provides a method of determining the presence of or predisposition of a NOVX-associated disorder in a subject. The method includes providing a nucleic acid sample, *e.g.*, RNA or DNA, or both, from the subject and measuring the amount of the NOVX nucleic acid in the subject nucleic acid sample. The amount of NOVX nucleic acid sample in the subject nucleic acid is then compared to the amount of a NOVX nucleic acid in a control sample. An alteration in the amount of NOVX nucleic acid in the sample relative to the amount of NOVX in the control sample indicates the subject has a NOVX-associated disorder.

In a still further aspect, the invention provides a method of treating or preventing or delaying a NOVX-associated disorder. The method includes administering to a subject in which such treatment or prevention or delay is desired a NOVX nucleic acid, a NOVX polypeptide, or a NOVX antibody in an amount sufficient to treat, prevent, or delay a NOVX-associated disorder in the subject.

Unless otherwise defined, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs. Although methods and materials similar or equivalent to those described herein can be used in the practice or testing of the present invention, suitable methods and materials are described below. All publications, patent applications, patents, and other references mentioned herein are incorporated by reference in their entirety. In the case of conflict, the present specification, including definitions, will control. In addition, the materials, methods, and examples are illustrative only and not intended to be limiting.

TABLE 1. Sequences and Corresponding SEQ ID Numbers

NOVX Assignment	Internal Identification	SEQ ID NO (nucleic acid)	SEQ ID NO (polypeptide)	Homology
1	27824582.0.105	1	2	Collagen family
2	27824582.0.50	3	4	Collagen family
3	CG51785-06	5	6	Collagen family
4	AC008687_A	7	8	Voltage-gated potassium channel
5	30412306.0.100	9	10	Tuftelin
6	30412306.1	11	12	Tuftelin
7	30412306.0.16	13	14	Tuftelin
8	AC005924_A	15	16	Neuronal antigen
9	h_hn0052k24_A	17	18	Fatty acid binding protein
10	h_hn0052k24_B	19	20	Fatty acid binding protein
11	AL096677_A	21	22	Cystatin

NOVX nucleic acids and their encoded polypeptides are useful in a variety of applications and contexts. The various NOVX nucleic acids and polypeptides according to the invention are useful as novel members of the protein families according to the presence of domains and sequence relatedness to previously described proteins. Additionally, NOVX nucleic acids and polypeptides can also be used to identify proteins that are members of the family to which the NOVX polypeptides belong.

For example, NOV1-3 are homologous to members of the collagen family of proteins that are important in determining cell shape and migration. Thus, the NOV1-3 nucleic acids and polypeptides, antibodies and related compounds according to the invention will be useful in therapeutic and diagnostic applications in disorders characterized by altered cell motility, proliferation and migration, *e.g.* cancer, angiogenesis and wound healing.

Also, NOV4 is homologous to members of the potassium channel family of proteins present in all eukaryotic cells which maintain membrane potential and modulate electrical excitability in neurons. Thus, the NOV4 nucleic acids and polypeptides, antibodies and related compounds according to the invention will be useful in therapeutic and diagnostic applications in neurological disorders, *e.g.* episodic ataxia, autosomal dominant myokymia, stroke, Parkinson's disease, and Alzheimer's disease.

Further, NOV5-7 are homologous to members of the tuftelin family of proteins that are important in enamel mineralization. Therefore, NOV5-7 nucleic acids and polypeptides, antibodies and related compounds according to the invention will be useful in therapeutic and

diagnostic applications in disorders characterized by enamel defects such as amelogenesis imperfecta and other disorders involving enamel defects, including hypoplasia and hypomineralization.

5 Still further, NOV8 is homologous to a family of neuronal antigen-like proteins that are important in paraneoplastic neurological disorders. Thus, NOV8 nucleic acids and polypeptides, antibodies and related compounds according to the invention will be useful in therapeutic and diagnostic applications in disorders characterized by paraneoplastic neurological disorders, *e.g.* paraneoplastic limbic of brain-stem encephalitis occurring during testicular cancer.

10 Also, NOV9-10 are homologous to a family of fatty acid-binding proteins important in keratinocyte differentiation. Thus NOV 9-10 nucleic acids and polypeptides, antibodies and related compounds according to the invention will be useful in therapeutic and diagnostic applications in disorders characterized by aberrant keratinocyte differentiation, *e.g.* lesional psoriatic skin.

15 Finally, NOV11 is homologous to a family of cystatin-like proteins that are important in protecting eukaryotic cells from inappropriate proteolysis. Thus, NOV11 nucleic acids and polypeptides, antibodies and related compounds according to the invention will be useful in therapeutic and diagnostic applications in disorders characterized by inappropriate proteolysis, *e.g.* atherosclerosis and abdominal aortic aneurysm.

20 The NOVX nucleic acids and polypeptides can also be used to screen for molecules, which inhibit or enhance NOVX activity or function. Specifically, the nucleic acids and polypeptides according to the invention may be used as targets for the identification of small molecules that modulate or inhibit, *e.g.*, neurogenesis, cell differentiation, cell motility, cell proliferation, hematopoiesis, wound healing and angiogenesis.

25 Additional utilities for the NOVX nucleic acids and polypeptides according to the invention are disclosed herein.

NOV1

A NOV1 sequence according to the invention is a nucleic acid sequence encoding a polypeptide related to the collagen family of proteins. A NOV1 nucleic acid is expressed in pancreas, salivary gland, lung and lung tumor. A NOV1 nucleic acid and its encoded polypeptide includes the sequences shown in Table 2. The disclosed nucleic acid (SEQ ID NO:1) is 1,949 nucleotides in length and contains an open reading frame (ORF) that begins with an ATG initiation codon at nucleotides 750-752 and ends with a TAG stop codon at nucleotides 1644-1646. The representative ORF encodes a 298 amino acid polypeptide (SEQ ID NO:2) with a predicted molecular weight of 30,567.2 daltons (Da). PSORT analysis of a NOV1 polypeptide predicts a cytoplasmic protein with a certainty of 0.4500. There is no apparent N-terminal signal sequence. Putative untranslated regions upstream and downstream of the coding sequence are underlined in SEQ ID NO: 1.

TABLE 2.

15 GTGTTGCCTCTTGAATGAAAAACAGAAACACCCAAGGCAAAATGGTAATGGCCTG
TCCACTGAAAAGCAGAAGCCCCACATGAGCAAGCTGCAGGCAGCTGGCAGGCACCG
ATTCTGCTGTCCTGTTTTGGATGCTATCTAACATCTTCATGTTCAACCCAGAGAAGA
AACATCCCGCCGTTGCCCTGGGGCCCTCTCATCCACAGCAGGTTTCGAGCCTTCCC
CAGCCCTCGGGATGGACAACCCTTGAGAAGCAGAGGTCAGGGAACCCTGACCCCGC
20 CACCCTTGCCAGGCCATCCGCTGCCCTCACAGGCACAGACAGAAGGCCTCTGTCCG
TGGCCAGGGCACTCCATGGGGAAGAAACAGGCCCTGTTCCCTCCCTGCTCACCCTT
CACCCAGCTCAGCTGGCACAATACTGCCACCACACCTTCACCCTGCCTAGCCCA
ACCTGGCAGGGCCTCGGAGTAGCCTGCCAGCTAAAATACGGGTTGCCAGATAACT
GTGAATGTCAGATAAGAATCTTCTGGGACGAGTATGTCCCATGCCATATTGGGACA
25 TACTTACACTAATAAATTTCTGTTTATCTGAACTCAAATTTGCCTGGGCGTCCTGTA
CTTTTCTTAATAAATTTGGTGCCTCTACACACAAGGTCCCTGGGGTGGGGGGGCAC
AGGAGCAAGCCCCTTCCCAGGCTGGGTCCCTGCCGGCATCTCCACAGGCCAGGAC
TGGCCACCCAGATGGAGCCCGTGCCAGGCAGCCGGCGACAGACGGACAAAGGCTG
CTCAGGAGACACTGCACACCTTCCTCTTTCTTGTCTGGGGGCTCAAGAATCCAGACG
30 CCCACCTCCCCGAGCGAGCACCAAGACAGGAAGCCAACCTGCAATGCCAGCCAC
TGCGACCACAGGGCTCTGCCGGGGTCTGCCGGAACCCAGGGTTCCGGTCCAGAAG
CCAGGGATAAATGCCGCTTCTCTATAGGGACAGTCAGAGTAGAGAGGGGGAGGCC
TACAGTCTCACCTGCAGGGAGAGGAAGTCCTCGGGGCGGGCACGTGGGGGGCCTGA
CAGCTCCGAGCACACCCGGCCACAGTGACCACGGACTGCACACGCAGAAGCAGTCT
35 GGATCCCACGCGTGGCTGTGCTGCCAGCAGACAGCACCAACCTCCCATGCTCCTCA
TCACAGGAAAAGAGACCAGCAGCATCTTGCCAGGCATGGTGGGGCCCCTCCGCCA
CAGCCTAGGAGTCCAGGCCACCCACCCTCACAGCACTGGAGTGCCTGGGTGAGTGA
GGCCCTGGGACGGGCCTGCGGGCACAGGGGGACAGAGGGTTCGGGGAGGGCGGCG
CAGCCCCACGAAGGGCTCCTCCCAAGCCTGTGTGGGGCCAGGGGAGCTGCACCTC
40 CGGGATGGGACAAGGCAGGGTCCTGGCTTTCATCAGCCACAGCACAGCTGCCACAG
GGCACAAGGACGGCTGAGAGACGAGGTCTCACCCACACCATGGGGAAACCGA

GGCATGGGAAGGTAGGGGGGGGCAGCCAGGCTGGCGAGATCACAGGCAGG
 CAGGCCTGAAGGCCGAGCAATGCAGCCACTAGGAAGGCATGAGTTGGGGTCGGGGT
 GTCCCCAGCCCTAGAGCCCAAAGCTGCCACCACTCCCCACCCCAACATGGGTGGG
 GGCAGGGAGAGCTCTTCTTGGGACCAATCCCCAAAACCATGCGCAGTGGGCCCCGGCT
 5 GGAGCCCAGGCAGCAGGCATCCTCTCTGCCAGGGTGAGAACTGGGCCCTCATGTC
AGGCTGGAAGGGGGGTCTCCAGGTGGGGAGAAAGAACAGGAAGGAACCAGGCCCC
TCCCTCGAGGGACCCCGCACCCAGGCTGCTCCCTGAGCGTGGGGTGGGCTCAGCGC
AATTGGGTCCAGACACCTGTCCCGGGCAGCCGTCTCGA (SEQ ID NO.: 1)

10 MEPVPGSRRQTDKGCSGDTAHLPLSCLGAQESRRPPPRASTKTGSQPAMPSPLRPQGS
 GVLPEPRVPVQKPGINAASPIGTVRVERGRPTVSPAGRGSPRGGHVGGTAPSTPGHSDH
 GLHTQKQSGSHAWLCCQQTAPNLPCSSSQEKRPASLPGMVGPLRHSLGVQATHPHST
 GVRGSVRPWDGPAGTGGQVRGRRSPTKGSSQACVGPRGAAPPGWDKAGSWLSSAT
 AQLPQGTKGRLRDEVLTHTMGKPRHGKVGGAARLAPRSQAGRPEGRAMQPLGRHEL
 15 GSGCPQP (SEQ ID NO.: 2)

A NOV1 polypeptide has homology (33% identity) with human collagen type I alpha
 (EMBL Accession No.: CAA67261), and has homology (31% identity) with *Strongylocentrotus*
purpuratus (Purple sea urchin) alpha-1 collagen (EMBL Accession No.: Q26634). A region of a
 20 NOV1 polypeptide also has a high degree of homology (100% identity) with the human
 polypeptide sequence ORF652 (ORFX; PatP Accession No.: B40888) as is shown in Table 38.

TABLE 38.

NOV1:	1	MEPVPGSRRQTDKGCSGDTAHLPLSCLGAQESRRPPPRASTKTGSQPAMPSPLRPQGSAG	60

ORFX:	1	MEPVPGSRRQTDKGCSGDTAHLPLSCLGAQESRRPPPRASTKTGSQPAMPSPLRPQGSAG	60
NOV1:	61	VLPEPRVPVQKPGINAASPIGTVRVERGRPTVSPAGRGSPRGGHVGGTAPSTPGHSDHG	120

ORFX:	61	VLPEPRVPVQKPGINAASPIGTVRVERGRPTVSPAGRGSPRGGHVGGTAPSTPGHSDHG	120
NOV1:	121	LHTQKQSGSHAW	132 (SEQ ID NO.: 71)

ORFX:	121	LHTQKQSGSHAW	132 (SEQ ID NO.: 72)
35	Where * indicates identity.		

NOV2

A NOV2 sequence according to the invention is a nucleic acid sequence encoding a
 polypeptide related to the collagen family of proteins. A NOV2 nucleic acid is expressed in
 40 pancreas, salivary gland, lung and lung tumor. A NOV2 nucleic acid and its encoded polypeptide
 includes the sequences shown in Table 3. The disclosed nucleic acid (SEQ ID NO:3) is 2,092
 nucleotides in length and contains an open reading frame (ORF) that begins with an ATG
 initiation codon at nucleotides 767-769 and ends with a TAG stop codon at nucleotides 1616-

1618. The represen ORF encodes a 283 amino acid polypeptide (SEQ ID NO:4) with a predicted molecular weight of 29,009.5 daltons (Da). PSORT analysis of a NOV2 polypeptide predicts a cytoplasmic protein with a certainty of 0.4500. There is no apparent N-terminal signal sequence. Putative untranslated regions upstream and downstream of the coding sequence are underlined in SEQ ID NO: 3.

TABLE 3.

10 TGCCCCGGGCAGGTGGGCGTGTTGCCTCTTGCAATGAAAAACAGAAACACCCAAGGC
AAAATGGTAATGGCCTGTCCACTGAAAAGCAGAAGCCCCACATGAGCAAGCTGCAG
GCAGCTGGCAGGCACCGATTCTGTCTGTCTGTTTTGGATGCTATCTAACATCTTCAT
GTTCAACCCAGAGAAGTTTCATCCCGCCGTTGCCCTGGGGCCCTCTCATCCCACAGC
AGGTTTCAAGCCTTCCCCAGCCCTCGGGATGGACAACCTTGAGAAGCAGAGGTCA
GGGAACCCTGACCCCGCCACCCTTGCCCAGGCCATCCGCTGCCCTCACAGGCACAG
15 ACAGAAGGCCTCTGTCCGTGGCCAGGGCACTCCATGGGGAAGAAACAGGCCCTGTT
CCCTCCCTGCTCACCCTTCACCCAGCTCAGCTGGCACAAAAATACTGCCACCACAC
CTTCACCCTGCCTAGCCCAACCTGGCAGGGCCTCGGAGTAGCCTGCCAGCTAAAATA
CGGGTTGCCCAGATAACTGTGAATGTCAGATAAGAATCTTCTGGGACGAGTATGTCC
CATGCCATATTTGGGACATACTTACACTAATAAATTTCTGTTTATCTGAAACTCAAAT
20 TTGCCTGGGCGTCTGTACTTTTCTTAACTAAATTTGGTGCCTCTACACACAAGGTCC
CTGGGGTGGGGGGGCACAGGAGCAAGCCCCTTCCCAGGCTGGGTCCCTGCCGGCAT
CTCCACAGGCCAGGACTGGCCACCCAGATGGAGCCCGTGCCAGGCAGCCGGCGAC
AGACGGACAAAGGCTGCTCAGGAGACACTGCACACCTTCTTCTTGTCTGGGGG
CTCAAGAATCCAGACGCCCACCTCCCCGAGCGAGCACCAAGACAGGAAGCCAACCT
25 GCAATGCCCAGCCCACTGCGACCACAGGGCTCTGCCGGGGTCTGCCGGAACCCAG
GGTTCGGGTCCAGAAGCCAGGGATAAATGCCGCTTCTCCTATAGGGACAGTCAGAG
TAGAGAGGGGGAGGCCTACAGTCTCACCTGCAGGGAGAGGAAGTCCTCGGGGCGGG
CACGTGGGGGGCCTGACAGCTCCGAGCACACCCGGCCACAGTGACCACGGACTGCA
CACGCAGAAGCAGTCTGGATCCCACGCGTGGCTGTGCTGCCAGCAGACAGCACCCA
30 ACCTCCCATGCTCCTCATCACAGGAAAAGAGACCAGCAGCATCTCTGCCAGGCATG
GTGGGGCCCCCTCCGCCACAGCCTAGGAGTCCAGGCCACCCACCCTCACAGCACTGG
AGTGCGTGGGTCAAGTGAAGGCCCTGGGACGGGCCTGCGGGCACAGGGGGACAGAGG
GTTCCGGGAGGGCGGCGCAGCCCCACGAAGGGCTCCTCCCAAGCCTGTGTGGGGCC
CAGGGGAGCTGCACCTCCGGGATGGGACAAGGCAGGGTCCTGGCTTTCATCAGCCA
35 CAGCACAGCTGCCACAGGGCACAAAAGGACGGCTGAGAGACGAGGTCTCACCCAC
ACCATGGGGAAACCGAGGCATGGGAAGGTTGGAGGGGGGGCAGCCAGGCTGGCGC
CAAGATCACAGGCAGGCAGGCCTGAAGGCCGAGCAATGTAGCCACTAGGAAGGCA
TGAGTTGGGGTTCGGGGTGTCCCCAGCCCTAGAGCCCAAAGCTGCCACCACTCCCCAC
CCCCAACATGGGTGGGGGCAGGGAGAGCTCTTCTTGGGACCAATCCCAAACCATG
40 CGCAGTGGGCCCCGGCTGGAGCCCAGGCAGCAGGCATCCTCTCTGCCAGGGTGAGAA
ACTGGGCCCTCATGTGAGGCTGGAAGGGGGGTCTCCAGGTGGGGAGAAAGAACAGG
AAGGAACCAGGCCCTCCCTCGAGGGACCCCGCACCCAGGCTGCTCCCTGAGCGTG
GGGTGGGCTCAGCGCACCTGGGTCCACACAGGGACCTGGCAAAGCTGTAGAGGCTG
TGGGAGGGGCTGCCGCTGGATGGGGTACAGGCCCGCCGCCCTTCTGAGAGGACAG

GGGAGGCCAGACGCTGATGCGGACTGACCGCCCATCAGACGGGATGTAG
AGGGCTCCCC (SEQ ID NO.: 3)

MEPVPGSRRQTDKGCSGDTAHLPLSCLGAQESRRPPPRASTKTGSQPAMPSPLRPQGS
GVLPEPRVPVQKPGINAASPIGTVRVERGRPTVSPAGRGSPRGGHVGGTLAPSTPGHSDH
GLHTQKQSGSHA WLCCQQTAPNLPCSSSQEKRPAA SLPGMVGPLRHSLGVQATHPHST
GVRGSVRPWDGPAGTGGQVRGGRSPTKGSSQACVGPRGAAPGWDKAGSWLSSAT
AQLPQGTGRLRDEVLTHTMGKPRHGKVGGGAARLAPRSQAGRPEGRAM (SEQ ID
NO.: 4)

A NOV2 polypeptide has homology (33% identity) with human collagen type I alpha
(EMBL Accession No.: CAA67261), and has homology (31% identity) with *Strongylocentrotus*
purpuratus (Purple sea urchin) alpha-1 collagen (EMBL Accession No.: Q26634). A region of a
NOV2 polypeptide also has a high degree of homology (100% identity) with the human
polypeptide sequence ORF652 (ORFX; PatP Accession No.: B40888).

NOV3

A NOV3 sequence according to the invention is a nucleic acid sequence encoding a
polypeptide related to the collagen family of proteins. NOV3 was identified as is described in
Example 2 and is present in at least lymphoid tissue, mammary gland/breast tissue, pancreas and
salivary gland. A NOV3 nucleic acid and its encoded polypeptide includes the sequences shown
in Table 4. The disclosed nucleic acid (SEQ ID NO:5) is 1,011 nucleotides in length and
contains an open reading frame (ORF) that begins with an ATG initiation codon at nucleotides
31-33 and ends with a TAG stop codon at nucleotides 925-927. The representative ORF encodes
a 298 amino acid polypeptide (SEQ ID NO: 6) with a predicted molecular weight of 29,009.5
daltons (Da). PSORT analysis of a NOV2 polypeptide predicts a cytoplasmic protein with a
certainty of 0.4500. There is no apparent N-terminal signal sequence. Putative untranslated
regions upstream and downstream of the coding sequence are underlined in SEQ ID NO: 5.

TABLE 4.

ATCTCCCACAGGCCAGGACTGGCCACCCAGATGGAGCCCGTGCCAGGCAGCCGGCG
ACAGACGGACAAAGGCTGCTCAGGAGACACTGCACACCTTCCTCTTTCTTGTCTGGG
GGCTCAAGAATCCAGACGCCCACCTCCCCGAGCGAGCACCAAGACAGGAAGCCAAC
CTGCAATGCCAGCCCACTGCGACCACAGGGCTCTGCCGGGGTCTGCGCGGAACCC
AGGGTTCCGGTCCAGAAGCCAGGGATAAATGCCGCTTCTCCTATAGGGACAGTCAA
GGTAGAGAGGGGGAGGCCTACAGTCTCACCTGCAGGGAGAGGAAGTCCTCGGGGCG

GGCACGTGGGGGGGACAGCTCCGAGCACACCCGGCCAGTGACCACGGACTG
CACACGCAGAAGCAGTCTGGATCCACGCGTGGCTGTGCTGCCAGCAGACAGCACC
CAACCTCCCATGCTCCTCATCACAGGAAAAGAGACCAGCAGCATCTCTGCCAGGCAT
5 GGTGGGGCCCCCTCCGCCACAGCCTAGGAGTCCAGGCCACCCACCCTCACAGCACTG
GAGTGCGTGGGTCAAGTGAAGGCCCTGGGACGGGCCTGCGGGCACAGGGGGACAGAG
GGTTCGGGGAGGGCGGCGCAGCCCCACGAAGGGCTCCTCCCAAGCCTGTGTGGGGC
CCAGGGGAGCTGCACCTCCGGGATGGGACAAGGCAGGGTCCTGGCTTTCATCAGCC
ACAGCACAGCTGCCACAGGGCACAAAAGGACGGCTGAGAGACGAGGTCCTCACCCA
CACCATGGGGAAACCGAGGCATGGGAAGGTTGGAGGGGGGGCAGCCAGGCTGGCG
10 CCAAGATCACAGGCAGGCAGGCCTGAAGGCCGAGCAATGCAGCCACTAGGAAGGC
ATGAGTTGGGGTCGGGGTGTCCCCAGCCCTAGAGCCCCAAAGCTGCCACCACTCCCC
ACCCCCAACATGGGTGGGGGCAGGGAGAGCTCTTCTTGGGACCAATCCCAAAACCA
TGCG (SEQ ID NO.: 5)

15 MEPVPGSRRQTDKGCSGDTAHLPLSCLGAQESRRPPPRASTKTGSQPAMPSPLRPQGSA
GVLPEPRVPVQKPGINAASPIGTVKVERGRPTVSPAGRGSPRGGHVGGTLAPSTPGHSDH
GLHTQKQSGSHAWLCCQQTAPNLPCSSSQEKRPAAASLPGMVGLRHSGLGVQATHPHST
GVRGSVRPWDGPAGTGGQVRGRRSPTKGSSQACVGPARGAAPPGWDKAGSWLSSAT
AQLPQGTGRLRDEVLTHTMGKPRHGKVGGAARLAPRSQAGRPEGRAMQPLGRHEL
20 GSGCPQP (SEQ ID NO.: 6)

One or more consensus positions (Cons. Pos.) of the NOV3 nucleotide sequence have been identified as SNPs as shown in Table 5. "Depth" represents the number of clones covering the region of the SNP. The Putative Allele Frequency (Putative Allele Freq.) is the fraction of all the clones containing the SNP. The sign ">" means "is changed to".

TABLE 5.

Cons.Pos.: 56 Depth: 59 Change: C > T Putative Allele Freq.: 0.034
Cons.Pos.: 83 Depth: 59 Change: G > A Putative Allele Freq.: 0.034
Cons.Pos.: 282 Depth: 52 Change: G > A Putative Allele Freq.: 0.038
30 Cons.Pos.: 323 Depth: 51 Change: G > A Putative Allele Freq.: 0.039
Cons.Pos.: 337 Depth: 52 Change: C > T Putative Allele Freq.: 0.038
Cons.Pos.: 432 Depth: 67 Change: G > A Putative Allele Freq.: 0.030
Cons.Pos.: 485 Depth: 69 Change: C > T Putative Allele Freq.: 0.029
Cons.Pos.: 746 Depth: 38 Change: A > G Putative Allele Freq.: 0.053
35

A NOV3 polypeptide has homology (33% identity) with human collagen type I alpha (EMBL Accession No.: CAA67261), and has homology (31% identity, 39% similarity) with *Strongylocentrotus purpuratus* (Purple sea urchin) alpha-1 collagen (COLa1; EMBL Accession No.: Q26634), as is shown in Table 6.). A region of a NOV3 polypeptide also has a high degree

of homology (99 % ty) with the human polypeptide sequence RF652 (ORFX; PatP Accession No.: B40888).

TABLE 6.

5	NOV3:	2	EPVPGSRRQTDKGCSDTAHLPLSCLGAQESRRPPPRASTKTGSQPAMPSPLRPQGSAGV	61
			** ** * + * ** * * + * * * + * + * *	
	COLa1:	751	EPGPAGA-PGQRGDPDGLG--PQGSPPSPGFAGPPGR-SGNPQPQEL-GPTGARGETGG	805
10	NOV3:	62	LPEPRVPVQKPGINAASPIGTVKVERGRPTVSPAGRGSPRG--GHVG-----GLTAP	111
			* * * ** ** + * * + * * * * * + * * *	
	COLa1:	806	-PGPSGPTGDPG--PQGPIGAPGQGERGETGPQGQGGPPGPIGSLGAPGAQGPPGPTGP	862
	NOV3:	112	S----TPGHSD-HGLHTQKQSGSHAWLCCQQTAPN--LPCSSSQEKRPAAASLPGMVGPLR	164
			* +** * * * * * +* + * + * * *	
15	COLa1:	863	SGNAGSPGQPGARGEPPQSGSPGQPGLA-GRTGPSGERGDKGNDGQSGPPGPPGPAGPAG	921
	NOV3:	165	HSLGVQATHPHSTGVRGSRVPWDGPAGTGGQVRGRRSPKSGSSQACVGPARGAAPPGWD	224
			* * + * * * * * * * * + * +** * * * * * *	
	COLa1:	922	QS-GILGLAGGS-GPRGPGGP-AGPPGAAGSRGPAGK-SGDRGSPGA-VGPAGNPAGE	976
20	NOV3:	225	KA--GSWLSSATAQLPQGTKGRLRDEVLTHTMGKP-RHGKVG--GGAARLAPRSQAGRPE	279
			(SEQ ID NO.: 23)	
			** + * ****+ * + * * * + * * ** + *	
25	COLa1:	977	NGMPGSDGNDG-APGPQGSERGEKDTGASGANGSPGAPGPIGAPGAAGASGPRGETGST-	1034
			(SEQ ID NO.: 24)	

Where * indicates identity and + indicates similarity.

NOV1-3 are new members of the collagen family of proteins. NOV1-3 have a high degree of homology between each other, as is shown in Table 7 and thus represent a new sub-family of the collagen family of proteins. The collagens are the major structural glycoproteins of connective tissues. A unique primary structure and a multiplicity of post-translational modification reactions are required for normal fibrillogenesis. The post-translational modifications include hydroxylation of prolyl and lysyl residues, glycosylation, folding of the molecule into triple-helical conformation, proteolytic conversion of precursor procollagen to collagen, and oxidative deamination of certain lysyl and hydroxylysyl residues. Any defect in the normal mechanisms responsible for the synthesis and secretion of collagen molecules or the deposition of these molecules into extracellular fibers could result in abnormal fibrillogenesis; such defects could result in a connective tissue disease. Recently, defects in the regulation of the types of collagen synthesized and in the enzymes involved in the post-translational modifications have been found in heritable diseases of connective tissue. Thus far, the primary heritable disorders of collagen metabolism in man include lysyl hydroxylase deficiency in Ehlers-Danlos

syndrome type VI, p-collagen peptidase deficiency in Ehlers-Danlos syndrome type VII, decreased synthesis of type III collagen in Ehlers-Danlos syndrome type IV, lysyl oxidase deficiency in S-linked cutis laxa and Ehlers-Danlos syndrome type V, and decreased synthesis of type I collagen in osteogenesis imperfecta (See PMID: 1448, UI: 76096101).

5 Distinct collagen subtypes are recognized by specific cell surface receptors. Two of the best known collagen receptors are members of the integrin family and are named alpha1beta1 and alpha2beta1. Integrin alpha1beta1 is abundant on smooth muscle cells, whereas the alpha2beta1 integrin is the major collagen receptor on epithelial cells and platelets. Many cell types, such as fibroblasts, osteoblasts, chondrocytes, endothelial cells, and lymphocytes may
10 concomitantly express both of the receptors. Furthermore, the two receptors are connected to distinct signaling pathways and their ligation may lead to opposite cellular responses. (See PMID: 10963992).

 Connective tissues maintain shape against external and internal stress. They are molecular
hierarchies in which fundamental building units come together in tiers of increasing complexity
15 and mutual interactions, based on information carried in the precursor molecules secreted by cells. The collagen fibril is the end product of well-understood self-aggregation controlled by its amino acid sequences, but the interfibrillar amorphous ground substance has not hitherto been seen as structured by analogous aggregations prescribed by the primary structures of the
20 characteristic glycosaminoglycans dissolved therein. Transmission electron microscopy with morphometry and stereology has demonstrated their existence in tissues. Nuclear magnetic resonance defined their secondary structures, rotary shadowing electron microscopy delineated their aggregates in vitro, and molecular dynamics simulations showed how the latter can spring from the former. The driving forces to aggregation are hydrophobic and hydrogen bonding, offset by electrostatic repulsion between polyanionic charges. The relative stabilities of the aggregates
25 are determined by this balance, and hence by the position and number of their charges, particularly the sulfate ester groups. Corneal stroma is a system of collagen fibrils, highly ordered to ensure transparency, in which glycosaminoglycan aggregates are suggested to determine the ordered spacing as yardsticks in a way that has parallels in all connective tissues. (See PMID: 1612287, UI: 92307242).

30 Recent biochemical and immunohistochemical studies have described several components of basement membranes including heparan sulfate proteoglycan, 2 high molecular

weight glycoproteins (fibronectin and laminin), and 2 collagen types (IV and V). These collagens have several properties which distinguish them from other types that are located in the interstitium: (a) type IV forms an amorphous, felt-like matrix, and neither IV nor V is found in large, cross-banded fibrils, (b) both have an increased content of hydrophobic amino acids, (c) the precursor (pro) forms are larger than those of interstitial collagens, (d) type IV contains interruptions within the triple helix, and e) both IV and V are resistant to human skin collagenase but are substrates for selected neutral proteases derived from mast cells, macrophages, and granulocytes. By immunofluorescence staining, type IV collagen has been localized to basement membranes at the dermal-epidermal junction, in capillaries, and beneath endothelial cells in larger vessels. Ultrastructurally it has been shown to be a specific component of the lamina densa. Type V collagen has been localized to the pericellular matrices of several cells types and may be specific for extramembranous structures which are closely associated with basal laminae. Other collagenous proteins have been described which may be associated with the extracellular matrix. One of these is secreted by endothelial cells in culture and by peptide mapping represents a novel collagen type. It is secreted under ascorbate-free conditions and is highly sensitive to proteolytic degradation. It has been proposed that a dynamic reciprocity exists between cells and their extracellular matrix which partially determines cell shape, biosynthesis, migration, and attachment. Examples of phenotypic modulation in several of these phenomena have been shown with endothelial cells grown on different substrates and isolated from different vascular environments. (See PMID: 7045245, UI: 82215350).

NOV1-3 represent a new subfamily of the collagen family. NOV1-2 can be used to detect pancreas, salivary gland, lung and lung tumor, and NOV3 can be used to detect at least lymphoid tissue, mammary gland/breast tissue, pancreas and salivary gland. NOV1-3 are useful in determining changes in expression of genes contained within the collagen protein family. NOV1-3 satisfy a need in the art by providing new diagnostic or therapeutic compositions useful in the treatment of disorders associated with alterations in the expression of members of collagen-associated proteins. NOV1-3 nucleic acids, polypeptides, antibodies, and other compositions of the present invention are useful in the treatment and/or diagnosis of a variety of diseases and pathologies, including by way of nonlimiting example, those involving pancreatic cancer, breast cancer, lymphoma, and other disorders characterized by alterations in cell shape, motility and differentiation, *e.g.* pathological angiogenesis, and wound healing.

TABLE 7.

NOV2	MEPVPGSRRTDKGCSGDTAHLPLSCLGAQESRRPPPRASTKTGSQPAMPSPLRPQGSAG	
NOV1	MEPVPGSRRTDKGCSGDTAHLPLSCLGAQESRRPPPRASTKTGSQPAMPSPLRPQGSAG	
NOV3	MEPVPGSRRTDKGCSGDTAHLPLSCLGAQESRRPPPRASTKTGSQPAMPSPLRPQGSAG	
5	*****	
NOV2	VLPEPRVPVQKPGINAASPIGTVRVERGRPTVSPAGRGSPRGHVGGTLTAPSTPGHSDHG	
NOV1	VLPEPRVPVQKPGINAASPIGTVRVERGRPTVSPAGRGSPRGHVGGTLTAPSTPGHSDHG	
NOV3	VLPEPRVPVQKPGINAASPIGTVKVERGRPTVSPAGRGSPRGHVGGTLTAPSTPGHSDHG	
10	***** : *****	
NOV2	LHTQKQSGSHAWLCCQQTAPNLPCSSSQEKRPAASLPGMVGPLRHS LGVQATHPHSTGVR	
NOV1	LHTQKQSGSHAWLCCQQTAPNLPCSSSQEKRPAASLPGMVGPLRHS LGVQATHPHSTGVR	
NOV3	LHTQKQSGSHAWLCCQQTAPNLPCSSSQEKRPAASLPGMVGPLRHS LGVQATHPHSTGVR	
15	*****	
NOV2	GSVRPWDGPAGTGGQVRVGGRRSPTKGSSQACVGPRGAAPPGWDKAGSWLSSATAQLPQG	
NOV1	GSVRPWDGPAGTGGQVRVGGRRSPTKGSSQACVGPRGAAPPGWDKAGSWLSSATAQLPQG	
NOV3	GSVRPWDGPAGTGGQVRVGGRRSPTKGSSQACVGPRGAAPPGWDKAGSWLSSATAQLPQG	
20	*****	
NOV2	TKGRLRDEVLTHTMGKPRHGKVGGAARLAPRSQAGRPEGRAM-----	(SEQ ID NO.: 4)
NOV1	TKGRLRDEVLTHTMGKPRHGKVGGAARLAPRSQAGRPEGRAMQPLGRHELGS GCPQP	(SEQ ID NO.: 2)
NOV3	TKGRLRDEVLTHTMGKPRHGKVGGAARLAPRSQAGRPEGRAMQPLGRHELGS GCPQP	(SEQ ID NO.: 6)
25	*****	

Where * indicates identity and : indicates strong similarity.

NOV4

30 A NOV4 sequence according to the invention is a nucleic acid sequence encoding a polypeptide related to the voltage-gated potassium channel-like protein family of proteins. A NOV4 nucleic acid is found on human chromosome 19. A NOV4 nucleic acid and its encoded polypeptide includes the sequences shown in Table 8. The disclosed nucleic acid (SEQ ID NO: 7) is 1,747 nucleotides in length and contains an open reading frame (ORF) that begins with an

35 ATG initiation codon at nucleotides 38-40 and ends with a TGA stop codon at nucleotides 1715-1717. The representative ORF encodes a 559 amino acid polypeptide (SEQ ID NO: 8) with a predicted molecular weight of 61,642.7 Da. PSORT analysis predicts that a NOV4 polypeptide is a plasma membrane protein (certainty 0.6000). Putative untranslated regions upstream and downstream of the coding sequence are underlined in SEQ ID NO: 7.

TABLE 8.

40 GAAGCCTGATTCTGACGAAACACACGCACACGGAAACATGGAGAGACGCAGGACA
GGATCCCGGCGGCAGAAAGGACGGAGAGAAAGGGGACCCCGGGACGGGAAAGGCGC
AGAGCAGGCGCGGGCGGCGGCGGCGGGGCAGGGCAGGGCGGGCGTCCCGGCA
GAGGGCGCGCGGTCGCCCTGTCGCCCTCCGCCCCGCGGGGTACAGTGCCCCCTCC
45 CTCGCGCCCTAGCCGCCCTGCCGGGCTATTTTACGCGCGGACACCGGACACCGGACA

5 GCGAGTATTTCTTCGACCGGCACCGGCCAGCTTCGACGCCGTGCTCTACTACTACC
 AGTCCGGTGGGCGGCTGCGGCGGCGGCGCACGTGCCGCTCGACGTCTTCTTGAA
 GAGGTGGCCTTCTACGGGCTGGGCGCGGCGGCCCTGGCACGCCTGCGCGAGGACGA
 GGGCTGCCCGGTGCCGCGCGAGCGCCCCCTGCCCGCGCGCCTTCGCCCGCCAGCT
 10 GTGGCTGCTTTTCGAGTTTCCCGAGAGCTCTCAGGCCGCGCGCGTGTCTGCCGTAGT
 CTCGGTGTGTCATCCTCGTCTCCATCGTCTTCTGCCTCGAGACGCTGCCTGAC
 TTCCGCGACGACCGCGACGGCACGGGGCTTGCTGCTGCAGCCGACGCCGGCCCGGT
 GTTCCCCGCTCCGCTGAATGGCTCCAGCCAAATGCCTGGAAATCCACCCCGCTGCC
 CTTCAATGACCCGTTCTTCGTGGTGGAGACGCTGTGTATTTGTTGGTTCTCCTTTGAG
 CTGCTGGTACGCCTCCTGGTCTGTCCAAGCAAGGCTATCTTCTTCAAGAACGTGATG
 15 AACCTCATCGATTTTGTGGCTATCCTTCCCTACTTTGTGGCACTGGGCACCGAGCTGG
 CCCGGCAGCGAGGGGTGGGCCAGCAGGCCATGTCACTGGCCATCCTGAGAGTCATC
 CGATTGGTGCCTGTCTTCCGCATCTTCAAGCTGTCCCGGCACTCAAAGGGCCTGCAA
 ATCTTGGGCCAGACGCTTCGGGCCTCCATGCGTGAGCTGGGCCTCCTCATCTTTTTC
 TCTTCATCGGTGTGGTCTCTTTTCCAGCGCCGTCTACTTTGCCGAAGTTGACCGGGT
 20 GGACTCCCATTTCACTAGCATCCCTGAGTCCTTCTGGTGGGCGGTAGTCACCATGAC
 TACAGTTGGCTATGGAGACATGGCACCCGTCCTGTGGGTGGCAAGATAGTGGGCT
 CTCTGTGTGCCATTGCGGGCGTGCTGACTATTTCCCTGCCAGTGCCCGTCATTGTCTC
 CAATTTCACTACTTTTATCACCGGGAGACAGAGGGCGAAGAGGCTGGGATGTTCA
 GCCATGTGGACATGCAGCCTTGTGGCCCACTGGAGGGCAAGGCCAATGGGGGGCTG
 25 GTGGACGGGGAGGTACCTGAGCTACCACCTCCACTCTGGGCACCCCCCAGGGAACA
 CCTGGTCACCGAAGTGTGAGGAACAGTTGAGGTCTGCAGGACCTCACAC (SEQ ID
 NO.: 7)
 30 MERRRTGSRRQKDGEKGDPTGKAQSRRGRRRRRGRAGRASRQRARGRPVALRPAGV
 TVPPPSRPSRPAGLFYARTPDTGHRAGAAVGATRRFAGRRGRCARHGAAPPAAPCGCCE
 RLVLNVAGLRFETRARTLGRFPDILLGDPARRGRFYDDARREYFFDRHRPSFDAVLYYY
 QSGGRLRRPAHVPLDVFLLEEVAFYGLGAAALARLREDEGCPVPPERPLPRRAFARQLWL
 LFEFPESQAARVLAVVSVLVILVSIVVFCLETLPDFRDDRDTGLAAAAAAGPVFPAPL
 NGSSQMPGNPRLPFNDPFFVETLCICWFSFELLVRLVCPSKAIFFKNVMNLIDFVAIL
 35 PYFVALGTELARQRGVQQAAMSLAILRVIRLVRFKLSRHSKGLQILGQTLRASMRE
 LGLLIFFLFIGVVLFSsavYFAEVDVRVDSHFTSIPEFVWAVVTMTTVGYGDMAPVTVG
 GKIVGSLCAIAGVLTISLPVPVIVSNFSYFYHRETEGEEAGMFSHVDMQPCPLEGKANG
 GLVDGEVPELPPPLWAPPREHLVTEV (SEQ ID NO.: 8)

40 A NOV4 nucleic acid has a high degree of homology (100% identity) with an
 uncharacterized region on human chromosome 19, including the clone CTB-60B18 (CHR19;
 GenBank Accession No.: AC008687), as is shown in Table 9. A NOV4 polypeptide has
 homology (83% identity, 85% similarity) with a voltage-gated potassium channel-like protein
 from mouse (VGPC; EMBL Accession No.: AAC23664), as is shown in Table 10. A NOV4

polypeptide also has homology (69% identity, 81% similarity) with a human voltage-gated potassium channel protein (HGK5; EMBL Accession No.: P22001), as is shown in Table 11.

TABLE 9.

5	NOV4:	901	gttccccgctccgctgaatggctccagccaaatgcctggaaatccaccccgctgcctt	960
	Chr19:	82187	gttccccgctccgctgaatggctccagccaaatgcctggaaatccaccccgctgcctt	82128
10	NOV4:	961	caatgaccgcttcttcgtggaggagacgctgtgtatttggtgttctcctttgagctgct	1020
	Chr19:	82127	caatgaccgcttcttcgtggaggagacgctgtgtatttggtgttctcctttgagctgct	82068
15	NOV4:	1021	ggtacgcctcctggtctgtccaagcaaggctatcttcttcaagaacgtgatgaacctcat	1080
	Chr19:	82067	ggtacgcctcctggtctgtccaagcaaggctatcttcttcaagaacgtgatgaacctcat	82008
20	NOV4:	1081	cgattttgtggctatccttccctactttgtggcactgggcaccgagctggcccgccagcg	1140
	Chr19:	82007	cgattttgtggctatccttccctactttgtggcactgggcaccgagctggcccgccagcg	81948
25	NOV4:	1141	aggggtggggccagcaggccatgtcactggccatcctgagagtcacccgattggtgcgtgt	1200
	Chr19:	81947	aggggtggggccagcaggccatgtcactggccatcctgagagtcacccgattggtgcgtgt	81888
30	NOV4:	1201	cttcgcgcatcttcaagctgtcccgccactcaaagggcctgcaaatcttgggcccagacgct	1260
	Chr19:	81887	cttcgcgcatcttcaagctgtcccgccactcaaagggcctgcaaatcttgggcccagacgct	81828
35	NOV4:	1261	tcgggcctccatgcgtgagctgggcctcctcatcttttctcttcatcggtgtggtcct	1320 (SEQ
	ID NO.:25)			
	Chr19:	81827	tcgggcctccatgcgtgagctgggcctcctcatcttttctcttcatcggtgtggtcct	81768 (SEQ
	ID NO.:26)			

TABLE 10.

40	NOV4:	21	TGKAQSRGRRRRRRGRAGRASRQRARGRPVALRPAGVTVPPPSRPSRPAGLFYAR---	TP 77
			* * * * * + * + * * * * * * * * * + * *	
	VGPC:	3	TRKAQEIHGKAP--GGSVSTGVGTAEAGAP---SPAGVTPPPPRPGRTFHAIFTRRHRT	57
45	NOV4:	78	DTGHRAGAAVGATRRFAGRRGCARHGAAVPAAPCGCCERLVLNVAGLRFETRARTLGRFP	137
			* *	
	VGPC:	58	DWG--GCGVGATRPFTGRPGCARHGATVPAA-LRCCERLVLNVAGLRFETRARTLGRFP	113
50	NOV4:	138	DTLLGDPARRGRFYDDARREYFFDRHRPSFDAVLYYYQSGGRLRRPAHVPLDVFLEEVAF	197
			* *	
	VGPC:	114	DTLLGDPVRRSRFYDGARAEYFFDRHRPSFDAVLYYYQSGGRLRRPAHVPLDVFLEEVSF	173
	NOV4:	198	YGLGAAALARLREDEGCPVPPERPLPRRAFARQLWLLFEFPESQAARVLAVVSVLVILV	257
			* *	
	VGPC:	174	YGLGRR-LARLREDEGCAVA-ERPLPP-PFARQLWLLFEFPESQAARVLAVVSVLVILV	230

Table 11

5
10
15
20
25
30
35
40
45
50
55

NOV4: 258 SIVVFCLETLPDFRDDRGTGLAAAAAGPVFPAPLNGSSQPPRLPFNDPFFVET 317

VGPC: 231 SIVVFCLETLPDFRDDRDEPLAPVAAATGSFLARLNGSSPMPGAPPRQPFNDPFFVET 290

NOV4: 318 LCICWFSFELLVRLVCPSKAIFFKVMNLIDFVAILPYFVALGTALARQRGVQQAMSL 377

VGPC: 291 LCICWFSFELLVHLVACPSKAVFFKVMNLIDFVAILPYFVALGTALARQRGVQQPAMSL 350

NOV4: 378 AILRVIRLVRFIRIFKLSRHSKGLQILGQTLRASMRELGLLIFFLFIGVVLFSsavYFAE 437

VGPC: 351 AILRVIRLVRFIRIFKLSRHSKGLQILGQTLRASMRELGLLIFFLFIGVVLFSsavYFAE 410

NOV4: 438 VDRVDSHFTSIPESFWWAVVTMTTVGYGDMAPVTVGKIVGSLCAIAGVLTISLPVPVIV 497

VGPC: 411 VDRVDTHFTSIPESFWWAVVTMTTVGYGDMAPVTVGKIVGSLCAIAGVLTISLPVPVIV 470

NOV4: 498 SNFSYFYHRETEGEEAGMFSHVDMQPCGPLEGKANGGLVDGEVPELPPPLWAPPREHLVT 557

VGPC: 471 SNFSYFYHRETEGEEAGMYSHVDTQPCGTLEGKANGGLVDSEVPELPPPLWPPAGKHMVT 530

NOV4: 558 EV 559 (SEQ ID NO.: 27)
**
VGPC: 531 EV 532 (SEQ ID NO.: 28)
Where * indicates identity and + indicates similarity.

TABLE 11.

NOV4: 95 GRRGCARHGAAPVAPCG---CC-ERLVNLVAGLRFETRARTLGRFPDTLLGDPARRGR 149
* ** *+ *+ * ** *+*****+ ** *+***** ** *

HGK5: 27 GGGGCDRYEPLPPSLPAAGEQDCCGERVVINISGLRFETQLKTLCQFPETLLGDPKRRMR 86

NOV4: 150 FYDDARREYFFDRHRPSFDAVLYYYQSGGRLRRPAHVPLDVFLEEVAFYGLGAAALARLR 209
+ * *****+*****+*** *+*+* ** *+ *

HGK5: 87 YFDPLRNEYFFDRNRPSFDAILYYYQSGGRIRRVNVPIDIFSEEIRFYQLGEEAMEKFR 146

NOV4: 210 EDEGCPVPPERPLPRRAFARQLWLLFEFPSSQAARVLAVSVLVILVSIVVFCLETLPD 269
***** *+*****+*** ** *+*****+*****+*****+

HGK5: 147 EDEGFLREEERPLPRRDFQRQVWLLFEYPSSGPARGIAIVSVLVILISIVIFCLETLPE 206

NOV4: 270 FRDDRGTGLAAAAAGPVFPAPLNGSSQMPGNPPRLPFNDPFFVETLCICWFSFELLV 729

HGK5: 207 FRDEKD---YPASTSQDSFEAA--GNSTSGSRAGASSFSDPFFVETLCIIWFSFELLV 260

NOV4: 330 RLLVCPSKAIFFKVMNLIDFVAILPYFVALGTALARQRGVQQAMSLAILRVIRLVRF 389
* ***** *+***** *+*****+ ***** *+*****

HGK5: 261 RFFACPSKATFSRNIMNLIDIVAIIPYFITLGTALAERQNGQQAMSLAILRVIRLVRF 320

NOV4: 390 RIFKLSRHSKGLQILGQTLRASMRELGLLIFFLFIGVVLFSsavYFAEVDVDSHFTSIP 449
*****+*****+*****+*****+*****+*****+*****+*****

HGK5: 321 RIFKLSRHSKGLQILGQTLKASMRELGLLIFFLFIGVILFSsavYFAEADDPTSGFSSIP 380

NOV4: 450 ESFWWAVVTMTTVGYGDMAPVTVGKIVGSLCAIAGVLTISLPVPVIVSNFSYFYHRETE 509
*+*****+*****+*****+*****+*****+*****+*****

HGK5: 381 DAFWWAVVTMTTVGYGDMHPVTIGGKIVGSLCAIAGVLTIALPVPVIVSNFNFYHRETE 440

NOV4: 510 GEEAG V 519

*** + **

HGK5: 441 GEEQSQYMHV 450

Where * indicates identity and + indicates similarity.

- 5 NOV4 represents a new member of a sub-class of voltage-gated potassium channels that includes members from several species (e.g. human, mouse and rat), as is shown by CLUSTALW analysis in Table 12.

TABLE 12.

10	mKV	MTTR-----KAQEIHG--KAPGGSVSTGVGTAEG---APSPAGVTPP
	NOV4	MERRRTGSRRQKDGEKGDPTGKAQSRGRRRRRGRAGRASRQRARGRPVALRPAGVTVP
	hKV	MT-----VVPGDHLLLEPEVADGG-----G-APP
	rKV	MT-----VVPGDHLLLEPEAAGGG-----GGDPP
		* * * *
15	mKV	PPRPGRFTFHAIFTRRHRTPDWGGCGVGATRPFTGRPGCARHGATVPAALR-----CCE
	NOV4	PPSRPSRPAGLFYARTPDTGHRAGAAVGATRRFAGRRGCARHGAAPVPAAPCG-----CCE
	hKV	Q-----GGCG-GG-----GCDRYEPLPPSLPAAGEQDCCGE
	rKV	Q-----GGCVSGG-----GCDRYEPLPPALPAAGEQDCCGE
		. * . * * * : . * : * *
20		
25	mKV	RLVLNVAGLRFETRARTLGRFPDILLGDPVRRSRFYDGARAEYFFDRHRPSFDAVLYYYQ
	NOV4	RLVLNVAGLRFETRARTLGRFPDILLGDPARRGRFYDDARREYFFDRHRPSFDAVLYYYQ
	hKV	RVVINISGLRFETQLKTLQCFPETLLGDPKRRMRYFDPLRNEYFFDRNRPSFDAILYYYQ
	rKV	RVVINISGLRFETQLKTLQCFPETLLGDPKRRMRYFDPLRNEYFFDRNRPSFDAILYYYQ
		*:
30	mKV	SGGRLRRPAHVPLDVFLEEVSYFGLG-RRLARLREDEGCABA-ERPLPP-PFARQLWLLF
	NOV4	SGGRLRRPAHVPLDVFLEEVAFYGLGAAALARLREDEGCPVPPERPLPRRAFARQLWLLF
	hKV	SGGRIIRPVNVPIDIFSEEIRFYQLGEEAMEKFREDEGFLREEERPLPRRDFQRQVWLLF
	rKV	SGGRIIRPVNVPIDIFSEEIRFYQLGEEAMEKFREDEGFLREEERPLPRRDFQRQVWLLF
		****:
35	mKV	EFPESSQAARVLAVSVLVILVSIIVVFCLETLPDFRDDRDPGLAPVAAATGSFLARLNG
	NOV4	EFPESSQAARVLAVSVLVILVSIIVVFCLETLPDFRDDRDTGLAAAAAGPVFPAPLNG
	hKV	EYPESSGPARGIAIVSVLVILISIVIFCLETLPDFRDEKD----YPASTSQDSFEAAGNS
	rKV	EYPESSRPARGIAIVSVLVILISIVIFCLETLPDFRDEKD----YPASPSQDVFEAANNS
		*:
40	mKV	SSPMGPAPRPQFPNDPFFVETLCICWFSFELLVHLVACPSKAVFFKNVMNLIDFVAILP
	NOV4	SSQMPGNPRLPFPNDPFFVETLCICWFSFELLVRLVLCPSKAIFKNVMNLIDFVAILP
	hKV	TSGSRAGAS--SFSDPFFVETLCIIWFSFELLVRFFACPSKATFSRNIMNLIDIVAIIP
	rKV	TSGASSGAS--SFSDPFFVETLCIIWFSFELLVRFFACPSKATFSRNIMNLIDIVAIIP
45		:* . . . * .***** *****:..***** * :*:*****:***:*
	mKV	YFVALGTELARQRGVGPAMSLAILRVIRLVRFRIKLSRHSGKLQILGQTLRASMREL
	NOV4	YFVALGTELARQRGVQQAMSLAILRVIRLVRFRIKLSRHSGKLQILGQTLRASMREL
	hKV	YFITLGTELAERQNGQQAMSLAILRVIRLVRFRIKLSRHSGKLQILGQTLKASMREL
50	rKV	YFITLGTELAERQNGQQAMSLAILRVIRLVRFRIKLSRHSGKLQILGQTLKASMREL
		::***:.* ** *****:*****:*****:*****:*****:*****:

mKV GLLIFFLFIGVVLFFYFAEVDVRVDTHFTSIPESFWWAVVTMTTVGYGVTVGKG
 NOV4 GLLIFFLFIGVVLFFSSAVYFAEVDVRVDSHFTSIPESFWWAVVTMTTVGYGDMAPVTVGKG
 hKV GLLIFFLFIGVILFSSAVYFAEADDPSTSGFSSIPDAFWWAVVTMTTVGYGDMHPVTIGGK
 rKV GLLIFFLFIGVILFSSAVYFAEADDPSSGFNSIPDAFWWAVVTMTTVGYGDMHPVTIGGK
 *****:*****.* : *.***:***** ****:***

mKV IVGSLCAIAGVLTISLPVPVIVSNFSYFYHRETEGEEAGMYSHVD-----T
 NOV4 IVGSLCAIAGVLTISLPVPVIVSNFSYFYHRETEGEEAGMFSHVD-----M
 hKV IVGSLCAIAGVLTIALPVPVIVSNFNFYHRETEGEEQSQYMHVGSQHLSSSAEELRKA
 rKV IVGSLCAIAGVLTIALPVPVIVSNFNFYHRETEGEEQAQYMHVGSQHLSSSAEELRKA
 *****:*****.***** . : **.

mKV QPCGTLEG-----KANGGLVDSEVPELLPPLWPPAG-----KHMVTEV (SEQ ID NO.:31)
 NOV4 QPCGPLEG-----KANGGLVDGEVPELPPPLWAPPR-----EHLVTEV (SEQ ID NO.:8)
 hKV RSNSTLSKSEYMVIEEGGMNHSAPFQTPFKTGNSTATCTNNPNPNSCVNIKKIFTDV (SEQ ID NO.:32)
 rKV RSNSTLSKSEYMVIEEGGMNHSAPFQTPFKTGNSTATCTNNPNPNSCVNIKKIFTDV (SEQ ID NO.:33)
 :. .*. :*: . . *: . . :*:.*

Where * indicates identity, : indicates strong similarity, and . indicates weak similarity. Rat
 potassium channel KV1.3 (rKV; Accession No.: A435310; human potassium channel KV1.3
 (hKV; Accession No.: P22001); mouse potassium channel KV1.7 (mKV; Accession No.:
 AAC23664).

Potassium channels represent the most complex class of voltage-gated ion channels from
 both functional and structural standpoints. Present in all eukaryotic cells, their diverse functions
 include maintaining membrane potential, regulating cell volume, and modulating electrical
 excitability in neurons. The delayed rectifier function of potassium channels allows nerve cells to
 efficiently repolarize following an action potential. In *Drosophila*, four sequence-related K⁺
 channel genes--Shaker, Shaw, Shab, and Shal--have been identified. Each has been shown to
 have a human homolog.

By PCR of genomic DNA with primers based on regions conserved between *Drosophila*
 Shaker and a mouse voltage-gated potassium channel, Ramaswami and co-workers (See
 Ramaswami *et al.*, 1990, *Mol. Cell. Neurosci.* 1:214) isolated fragments of several related human
 genes. They used the fragments to screen cDNA libraries and cloned cDNAs encoding several
 potassium channels that they designated HuKI (KCNA1), HuKII (KCNA4; 176266), HuKIV
 (KCNA2; 176262), and HuKV (KCNA6; 176257). Like other Shaker-class potassium channels,
 the predicted 495-amino acid KCNA1 protein contains six hydrophobic segments, a positively
 charged region called S4 between hydrophobic segments 3 and 4, and a leucine zipper. KCNA1
 shares 98% amino acid identity with its rat homolog, RCK1. When expressed in *Xenopus*
 oocytes, KCNA1, KCNA4, and KCNA2 exhibited different voltage dependence, kinetics, and
 sensitivity to pharmacologic potassium channel blockers. KCNA1 and KCNA2 were

noninactivating channels and resembled delayed rectifiers, while KCNA4 was rapidly inactivating.

Chandy and colleagues (See Chandy *et al.*, 1990, Science 247:973) demonstrated that three closely related potassium channel genes, MK1, MK2, and MK3, are located at separate sites in the genome of the mouse. These genes, encoding subunits of voltage-dependent K⁺ channels, are homologous to the *Drosophila* Shaker gene. Curran *et al.* (See Curran *et al.*, 1992, Genomics 12:729) mapped the KCNA1 gene to chromosome 12 by use of human-rodent somatic cell panels and narrowed the localization to the distal short arm by in situ hybridization. Linkage studies had shown a maximum lod score of 2.72 at a recombination fraction of 0.05 between KCNA1 and the von Willebrand locus (VWF; 193400). Using interspecific backcrosses between *Mus musculus* and, Klocke *et al.* (See Klocke *et al.*, 1993, Genomics 18:568) mapped the Kcna1, Kcna5 (176267), and Kcna6 genes to mouse chromosome 6, close to the homolog of TP11 (190450), which is located on 12p13 in the human. Albrecht and co-workers (See Albrecht *et al.*, 1995, Receptors Channels 3:213) determined that a 300-kb cluster on chromosome 12p13 contains the human KCNA6, KCNA1, and KCNA5 genes arranged in tandem.

Browne *et al.* (See Browne *et al.*, 1994, Nature Genetics 8:136) performed mutation analysis of the KCNA1 coding region in four families with myokymia (rippling of muscles) with episodic ataxia, also known as episodic ataxia type 1 (EA1; 160120). They found four different missense mutations present in heterozygous state. For a comprehensive review of episodic ataxia type 1 and its causative mutations, (See Brandt and Strupp, 1997, Audiol. Neurotol. 2:373). Adelman *et al.* (See Adelman *et al.*, 1995, Neuron 15:1449) injected *Xenopus* oocytes with cDNAs corresponding to six different mutations associated with autosomal dominant myokymia with episodic ataxia. They demonstrated that coassembly of one or more episodic ataxia subunits with a wild type subunit can alter channel function, giving a dominant-negative effect.

NOV4 is a new member of the voltage-gated potassium channel-like protein family of proteins. It is useful as a marker for human chromosome 19. As a member of the voltage-gated potassium channel-like family of proteins, NOV4 nucleic acids, proteins, antibodies and other compositions of the present invention are useful in potential therapeutic applications implicated in Episodic Ataxia, type 1, Long QT Syndrome 1 and 2, Benign Neonatal Epilepsy, Jervell and Lange-Neilsen syndrome, Autosomal dominant deafness (DFNA 2), non-insulin dependent diabetes mellitus, CNS disorders, arrhythmia, seizure, asthma, hypertension therapy and/or other

pathologies and disorders. NOV4 may be used in drug screening or identification of therapeutics which modulate the channel and, therefore, modulate insulin secretion. Selective antagonists increase insulin release and thereby reduce hyperglycaemia associated with non-insulin-dependent diabetes mellitus.

5

NOV5

A NOV5 sequence according to the invention is a nucleic acid sequence encoding a polypeptide related to the tuftelin-like protein family of proteins. A NOV5 nucleic acid is expressed in fetal liver, testis, fetal lung, and thyroid gland. A NOV5 nucleic acid and its encoded polypeptide includes the sequences shown in Table 13. The disclosed nucleic acid (SEQ ID NO: 9) is 1,080 nucleotides in length and contains an open reading frame (ORF) that begins with an ATG initiation codon at nucleotides 45-47 and ends with a TGA stop codon at nucleotides 799-801. The representative ORF encodes a 251 amino acid polypeptide (SEQ ID NO: 10) with a predicted molecular weight of 29,229.5 Da. PSORT analysis of NOV5 predicts a cytoplasmic protein (certainty:0.4500), and NOV5 appears to lack an N-terminal signal sequence. Putative untranslated regions upstream and downstream of the coding sequence are underlined in SEQ ID NO: 9.

TABLE 13.

20 GTTTAAAGATGAAATGAGACATGACAGTACAAATCACAACTAGATGCAAAGTTTG
GATTTGCTTATGAAAAAGATAAAAGGAAAAGACCTACAGCTCTTAGAAATGAACAA
AGAGAATGAAGTATTGAAAATCAAGCTGCAAGCCTCCAGAGAAGCAGGAGCAGCA
GCTCTGAGAAACGTGGCCCAGAGATTATTTGAAAACCTACCAAACGCAATCTGAAGA
AGTGAGAAAGAAGCAGGAGGGCAGTAAACAATTACTCCAGGTTAACAAGCTTGAAA
25 AAGAACAGAAATTGAAACAACATGTTGAAAATCTGAATCAAGTTGCTGAAAAACTT
GAAGAAAAACACAGTCAAATTACAGAATTGGAGAACCTTGACAGAGAATGGAAAA
GGAAAAGAGAACACTACTAGAAAGAAAAGTGTCTTTGGAAAACAAGCTACTGCAAC
TCAAATCCAGTGCTACATATGGAAAAAGTTGCCAGGATCTTCAGAGGGAGATTTCCA
TTCTCCAGGAGCAGATCTCTCATCTGCAGTTTGTGATTCACTCCCAACATCAGAACCT
30 GCGCAGTGTCATCCAGGAGATGGAAGGATTAATAATTTAAAAGAACAAGACA
AAAGAATTGAAAATCTCAGAGAAAAGGTTAACATACTTGAAGCCCAGAATAAAGAA
CTAAAAACCCAGGTAGCACTTTTCATCTGAAACTCCTAGGACAAAGGTATCTAAGGCT
GTCTCTACAAGTGAATTGAAGACCGAAGGTGTTTCCCCTTATTTAATGTTGATTAGGT
TACGGAAATGAACTGGCTGGATGAAGATCTGATTAGAAAGACTGCGTGAGTCTTA
35 TTTATTCTCTGAAACACAGCCCAAGTTTCATGTTAAATGGCAAAATGCCATTATTT
AAATGGAACTTATTACATACCAATGGCTTTGCAAGAAGATGACATTTAGAAAAATCA
AACAAATCTATATTTAATGGATGGACTCTTCAAACTTACCAAATAGTTGAAGAAAC
CAGGTGCCTTCTCATGATGGAAGACAGATTCTGCTTTAAATTAATAAAAAAAAAAATC
TGAAAAA (SEQ ID NO.: 9)

5 MQSLDLLMKKIKGKDLQLLEMNKENEVLKIKLQASREAGAAALRNVAQRLFENYQTQS
 EEVRKKQEGSKQLLQVNKLEKEQKLKQHVENLNQVAEKLEEKHSQITELENLVQRMEK
 EKRTLLEKRLSLENKLLQLKSSATYGKSCQDLQREISILQEQISHLQFVIHSQHQNLRSVI
 QEMEGLKNNLKEQDKRIENLREKVNILEAQNKELKTQVALSSETPRTKVSKAVSTSELK
 TEGVSPYLMLRLRK (SEQ ID NO.: 10)

10 A NOV5 nucleic acid has a high degree of homology (99% identity) with a human
 cutaneous T-cell lymphoma-associated antigen se57-1 mRNA (CTCL; GenBank Accession No.:
 AF273051), as is shown in Table 14. A NOV5 polypeptide has homology (26% identity, 55%
 similarity) with a *bos taurus* tuftelin-like protein (bTUF; EMBL Accession No.: O97683), as is
 shown in Table 15.

TABLE 14.

15 NOV5: 51 agtttggatttgcttatgaaaaagataaaaggaaaagacctacagctcttagaaatgaac 110
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 CTCL: 432 agtttggatttgcttatgaaaaagataaaaggaaaagacctacagctcttagaaatgaac 491
 20 NOV5: 111 aaagagaatgaagtattgaaaatcaagctgcaagcctccagagaagcaggagcagcagct 170
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 CTCL: 492 aaagagaatgaagtattgaaaatcaagctgcaagcctccagagaagcaggagcagcagct 551
 25 NOV5: 171 ctgagaaacgtggcccagagattatttgaaaactaccaaacgcaatctgaagaagtgaga 230
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 CTCL: 552 ctgagaaacgtggcccagagattatttgaaaactaccaaacgcaatctgaagaagtgaga 611
 30 NOV5: 231 aagaagcaggagggcagtaacaattactccaggttaacaagcttgaaaaagaacagaaa 290
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 CTCL: 612 aagaagcaggaggacagtaacaattactccaggttaacaagcttgaaaaagaacagaaa 671
 35 NOV5: 291 ttgaaacaacatgttgaaaatctgaatcaagttgctgaaaaacttgagaaaaacacagt 350
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 CTCL: 672 ttgaaacaacatgttgaaaatctgaatcaagttgctgaaaaacttgagaaaaacacagt 731
 NOV5: 351 caaattacagaattggagaaccttgtagagagaatggaaaaggaaaagagaacactacta 410
 (SEQ ID NO.: 34)
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 40 CTCL: 732 caaattacagaattggagaaccttgtagagagaatggaaaaggaaaagagaacactacta 791
 (SEQ ID NO.: 35)

TABLE 15.

45 NOV5: 42 ALRNVAQRLFENYQTQSEEVRRKKQEGSKQLLQVNKLEKEQKLKQHVENLNQVAEKLEEKH 101
 *** * * * * ++ + * + +++ ** * + ++ ***
 BTUF: 171 SLRKTVDLLVKLQ----EAEQQHQSDCSAFKVTLSQYQREAKQSQVALQRAEDRAEQKE 226

AGGTCCTTTTCCTTTATCTTTTTCATAAGCAAATCCAACTCCAAACAAGAAGGATCC
ATTTCAGAATCAGAGCCCTGTTGAAGGTTTCCACAGTGCTTTGCATCTAGTTTGTGAT
TTGTACTGTCATGTCTTATTTTCATCTTTAAACATCTGGGTCCTGATCTTTTGCAGAGT
AGTTTCGAATCTTTTTCACATACTCGGTTTCTTCAATAATGTGAGCGGACGTAGACTCA
5 TACAAGGCAGAATTATCTTCCATCTTATCCCTTGGGGGAATTTCTGTGGTCACTGCCA
CTGTTGTCATTGTGAATTCTGGCCAAGACGAAGTAAAATTAATAGAGCTAAAACGCC
AACCTTGGTCTTTTAGAAGTTCAGAGATGTTTCCATCATATTAAGACTGGCTTCCCTC
TTCAACAAGGACCCTTTTACAGGAAATGTCTTGATGCCAGGAACTCCACTGGGGAA
GCCGCTGGAAAGGCACCTGGACACCCACACAC (SEQ ID NO.: 11)

10 GTGTGTGGGTGTCCAGGTGCCCTTCCAGCGGCTTCCCCAGTGGAGTTCCTGGCATCA
AGGACATTTCTGTAAAAGGGTCCTTGTGGAAGAGGGAAGCCAGTCTTAATATGATG
GAAACATCTCTGAACTTCTAAAAGACCAAGGTTGGCGTTTTAGCTCTATTAATTTTA
CTTCGCTTGGCCAGAATTCACAATGACAACAGTGGCAGTGACCACAGAAATCCCC
15 CAAGGGATAAGATGGAAGATAATTCTGCCTTGTATGAGTCTACGTCCGCTCACATTA
TTGAAGAAACCGAGTATGTGAAAAAGATTGCAACTACTCTGCAAAAGATCAGGACC
CAGATGTTTAAAGATGAAATAAGACATGACAGTACAAATCACAACTAGATGCAAA
GCACTGTGGAAACCTTCAACAGGGCTCTGATTCTGAAATGGATCCTTCTTGTGTCAG
TTTGGATTTGCTTATGAAAAAGATAAAAAGGAAAAGACCTACAGCTCTTAGAAATGA
20 ACAAAGAGAATGAAGTATTGAAAATCAAGCTGCAAGCCTCCAGAGAAGCAGGAGC
AGCAGCTCTGAGAAACGTGGCCCAGAGATTATTTGAAAACCTACCAAACGCAATCTG
AAGAAGTGAGAAAGAAGCAGGAGGACAGTAAACAATTACTCCAGGTAAACAAGCTT
GAAAAAGAACAGAAATTGAAACAACATGTTGAAAATCTGAATCAAGTTGCTGAAAA
ACTTGAAGAAAAACACAGTCAAATTACAGAATTGGAGAACCTTGTACAGAGAATGG
25 AAAAGGAAAAGAGAACACTACTAGAAAGAAAACCTGTCTTTGGAAAACAAGCTACTG
CAACTCAAATCCAGTGCTACATATGGAAAAAGTTGCCAGGATCTTCAGAGGGAGAT
TTCCATTCTCCAGGAGCAGATCTCTCATCTGCAGTTTGTGATTCACTCCCAACATCAG
AACCTGCGCAGTGTCTATCCAGGAGATGGAAGGATTAAAAAATAATTTAAAGAACA
AGACAAAAGAATTGAAAATCTCAGAGAAAAGGTTAACATACTTGAAGCCCAGAATA
30 AAGAACTAAAAACCCAGGTAGCACTTTCATCTGAAACTCCTAGGACAAAGGTATCT
AAGGCTGTCTCTACAAGTGAATTGAAGACCGAAGGTGTTTCCCCTTATTTAATGTTG
ATTAGGTTACGGAATGAACTGGCTGGATGAAGATCTGATTTAGAAAGACTGCGTG
AGTCTTATTTATTCTCTGAAACACAGCCCAAGTTTCATGTTAAAATGGCAAAATGCC
ATTATTTAAATGGAACCTTATTACATACCAATGGCTTTGCAAGAAGATGACATTTAG
35 AAAATCAAACAAATCTATATTTAATGGATGGACTCTTCAAACTTACCAAATAGTTG
AAGAAACCAGGTGCCTTCTCATGATGGAAGACAGATTCTGCTTTAAATTAACAAAAA
AAAAAATCTGAAAAA (SEQ ID NO.: 69)

MTTVAVTTEIPPRDKMEDNSALYESTSAHIEETEYVKKIRTTLQKIRTQMFKDEIRHDST
40 NHKLDKHKCNLQQGSDSEMDPSCSLDLLMKKIKGKDLQLLEMNKENEVLKIKLQAS
REAGAAALRNVAQQLFENYQTQSEVRKKQEDSKQLLQVNKLEKEQKLKQHVNLNQ
VAEKLEEKHSQITELNLVQRMEKEKRTLLERKLSLENKLLQLKSSATYGKSCQDLQREI
SILQEQISHLQFVIHSQHQNLRSVIQEMEGLKNNLKEQDKRIENLREKVNILEAQNKELKT
45 QVALSSETPRTKVSKAVSTSELKTEGVSPYLMILRLRK (SEQ ID NO.: 12)

bioRxiv preprint doi: <https://doi.org/10.1101/031301>; this version posted March 1, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

A NOV6 nucleic acid has a high degree of homology (99% identity) with a human cutaneous T-cell lymphoma-associated antigen se57-1 mRNA (CTCL; GenBank Accession No.: AF273051). A NOV6 polypeptide has homology (25% identity, 53% similarity) with a *bos taurus* tuftelin-like protein (bTUF; EMBL Accession No.: O97683).

5

NOV7

A NOV7 sequence according to the invention is a nucleic acid sequence encoding a polypeptide related to the tuftelin-like protein family of proteins. A NOV7 nucleic acid is expressed in fetal liver, fetal lung, testis and thyroid gland. A NOV7 nucleic acid and its encoded polypeptide includes the sequences shown in Table 17. The disclosed nucleic acid (SEQ ID NO: 13) is 1,442 nucleotides in length and contains an open reading frame (ORF) that begins with an ATG initiation codon at nucleotides 155-157 and ends with a TGA stop codon at nucleotides 1159-1161. The representative ORF encodes a 335 amino acid polypeptide (SEQ ID NO: 14) with a predicted molecular weight of 42,276.8 Da. PSORT analysis of NOV7 predicts a cytoplasmic protein (certainty:0.4500), and NOV7 appears to lack an N-terminal signal sequence. Putative untranslated regions upstream and downstream of the coding sequence are underlined in SEQ ID NO: 13.

TABLE 17.

GCCCAGGGGAGGAGCAGCACCGGGACCCCGCGTCGGCTGGGCGCCCCACAAGGGA
AGCCAGTCTTAATATGATGGAAACATCTCTGAACTTCTAAAAGACCAAGGTTGGCGT
TTTAGCTCTATTAATTTTACTTCGTCTTGGCCAGAATTCACAATGACAACAGTGACA
GTGACCACAGAAATCCCCCAAGGGATAAGATGGAAGATAATTCTGCCTTGTATGA
GTCTACGTCCGCTCACATTATTGAAGAAACCGAGTATGTGAAAAAGATTGCAACTAC
TCTGCAAAAGATCAGGACCCAGATGTTTAAAGATGAAATAAGACATGACAGTACAA
25 ATCACAAACTAGATGCAAAGCACTGTGGAAACCTTCAACAGGGGCTCTGATTCTGAA
ATGGATCCTTCTTGTTCAGTTTGGATTTGCTTATGAAAAAGATAAAAGGAAAAGAC
CTACAGCTCTTAGAAATGAACAAAGAGAATGAAGTATTGAAAATCAAGCTGCAAGC
CTCCAGAGAAGCAGGAGCAGCAGCTCTGAGAAACGTGGCCCAGAGATTATTTGAAA
ACTACCAAACGCAATCTGAAGAAAGTGAAGAAAGCAGGAGGACAGTAAACAATT
30 ACTCCAGGTTAACAAGCTTGAAAAAGAACAGAAATTGAAACAACATGTTGAAAATC
TGAATCAAGTTGCTGAAAACTTGAAGAAAAACACAGTCAAATTACAGAATTGGAG
AACCTTGTACAGAGAATGGAAAAGGAAAAGAGAACTACTAGAAAGAAAAGTGT
CTTTGGAAAACAAGCTACTGCAACTCAAATCCAGTGCTACATATGGAAAAAGTTGCC
AGGATCTTCAGAGGGAGATTTCATTCTCCAGGAGCAGATCTCTCATCTGCAGTTTG
35 TGATTCACTCCCAACATCAGAACCTGCGCAGTGTCATCCAGGAGATGGAAGGATTA
AAAAATAATTTAAAAGAACAAGACAAAAGAATTGAAAATCTCAGAGAAAAGGTTA
ACATACTTGAAGCCCAGAATAAAGAACTAAAAACCCAGGTAGCACTTTCATCTGAA
ACTCCTAGGACAAAGGTATCTAAGGCTGTCTCTACAAGTGAATTGAAGACCGAAGG
TGTTTCCCCTTATTTAATGTTGATTAGGTTACGGAAATGAACTGGCTGGATGAAGAT

CTGATTTAGAAAGAGCGTGAGTCTTATTTATTCTCTGAAACAGCCCAAGTTTC
 ATGTTAAAATGGCAAAATGCCATTATTTAAATGGAAGTTATTACATACCAATGGCTT
 TGCAAGAAGATGACATTTTCAGAAAAATCAAACAAATCTATATTTAATGGATGGACTCT
 TCAAACTTACCAAATAGTTGAAGAAACCAGGTGCCTTCTCATGATGGAAGACAGA
 5 TTCTGCTTTAAATTAATAAAAAAAAAAATCTGAAAAA (SEQ ID NO.: 13)

MTTVTVTTEIPPRDKMEDNSALYESTSAHIIETEYVKKIRTTLQKIRTQMFKDEIRHDS
 NHKLDKHCNGLQQGSDSEMDPSCCSLDLLMKKIKGKDLQLEMNKENEVLKIKLQAS
 REAGAAALRNVAQRLFENYQTQSEEVKKQEDSKQLQVKNLEKEQKLKQHVNLNQ
 10 VAEKLEEKHSQITELNLRVQMEKEKRTLLERKLSLENKLLQLKSSATYGKSCQDLQREI
 SILQEQISHLQFVIHSQHNLRSVIQEMEGLKNNLKEQDKRIENLREKVNILEAQNKELKT
 QVALSSETPRTKVSKAVSTSELKTEGVSPYMLIRLRK (SEQ ID NO.: 14)

A NOV7 nucleic acid has a high degree of homology (99% identity) with a human
 15 cutaneous T-cell lymphoma-associated antigen se57-1 mRNA (CTCL; GenBank Accession No.:
 AF273051). A NOV7 polypeptide has homology (26% identity, 55% similarity) with a *bos*
taurus tuftelin-like protein (bTUF; EMBL Accession No.: O97683).

NOV5-7 polypeptides have a high degree of homology between each other, as is shown
 20 in Table 18. NOV5-7 therefore represent a novel sub-family of the tuftelin-like protein family.

TABLE 18.

25	NOV7	MTTVTVTTEIPPRDKMEDNSALYESTSAHIIETEYVKKIRTTLQKIRTQMFKDEIRHDS
	NOV6	MTTVAVTTEIPPRDKMEDNSALYESTSAHIIETEYVKKIRTTLQKIRTQMFKDEIRHDS
	NOV5	-----
30	NOV7	TNHKLDKHCNGLQQGSDSEMDPSCCSLDLLMKKIKGKDLQLEMNKENEVLKIKLQASR
	NOV6	TNHKLDKHCNGLQQGSDSEMDPSCCSLDLLMKKIKGKDLQLEMNKENEVLKIKLQASR
	NOV5	-----MQSLDLLMKKIKGKDLQLEMNKENEVLKIKLQASR

35	NOV7	EAGAAALRNVAQRLFENYQTQSEEVKKQEDSKQLQVKNLEKEQKLKQHVNLNQVAEK
	NOV6	EAGAAALRNVAQRLFENYQTQSEEVKKQEDSKQLQVKNLEKEQKLKQHVNLNQVAEK
	NOV5	EAGAAALRNVAQRLFENYQTQSEEVKKQEGSKQLQVKNLEKEQKLKQHVNLNQVAEK

40	NOV7	LEEKHSQITELNLRVQMEKEKRTLLERKLSLENKLLQLKSSATYGKSCQDLQREISILQ
	NOV6	LEEKHSQITELNLRVQMEKEKRTLLERKLSLENKLLQLKSSATYGKSCQDLQREISILQ
	NOV5	LEEKHSQITELNLRVQMEKEKRTLLERKLSLENKLLQLKSSATYGKSCQDLQREISILQ

45	NOV7	EQISHLQFVIHSQHNLRSVIQEMEGLKNNLKEQDKRIENLREKVNILEAQNKELKTQVA
	NOV6	EQISHLQFVIHSQHNLRSVIQEMEGLKNNLKEQDKRIENLREKVNILEAQNKELKTQVA
	NOV5	EQISHLQFVIHSQHNLRSVIQEMEGLKNNLKEQDKRIENLREKVNILEAQNKELKTQVA

NOV7 LSSETPRTKVSKESELKTEGVSPYMLIRLRK (SEQ ID NO. :
 NOV6 LSSETPRTKVSKEAVSTSELKTEGVSPYMLIRLRK (SEQ ID NO. :12)
 NOV5 LSSETPRTKVSKEAVSTSELKTEGVSPYMLIRLRK (SEQ ID NO. :10)

5 Where * indicates identity and . indicates weak similarity.

NOV5-7 have similarity to tuftelin, a protein of the enamelin-family. Tuftelin is a novel acidic enamel protein thought to play a major role in enamel mineralization, and is involved in the etiology of autosomally inherited amelogenesis imperfecta (AI). AI is a diverse group of hereditary disorders characterized by a variety of developmental enamel defects including hypoplasia and hypomineralization.

Tuftelin is a novel acidic enamel protein thought to play a major role in enamel mineralization. Its identity and localization has been confirmed by amino acid composition, enzyme-linked immunosorbant assay, Western blots, indirect immunohistochemistry and high resolution protein-A gold immunocytochemistry. The deduced tuftelin protein (pI 5.2) contains 389 amino acids and has a calculated peptide molecular mass of 43,814 Da. Immunological studies suggest conservation of tuftelin structure between species throughout vertebrate evolution. The cDNA sequence encodes for several putative post-translation sites including one N-glycosylation consensus site, seven O-glycosylation sites and seven phosphorylation sites, as well as an EF-hand calcium-binding domain (with mismatch), localized towards the N-terminal region. At the C-terminal region (residues 252-345) tuftelin contains structurally relevant determinants for self assembly. Employing fluorescent *in situ* hybridization, the human tuftelin gene was mapped to chromosome 1q 21-31. Localization of the human tuftelin gene to a well-defined cytogenetic region may be important in understanding the aetiology of autosomally inherited amelogenesis imperfecta, the most common enamel hereditary disease. (See Deutsch *et al.*, 1997, Ciba Found Symp 205:135-47; discussion 147-155).

The bovine tuftelin gene has been compared to that of bovine tuftelin cDNA. The analyses demonstrated that the cDNA contains a 1014-bp open reading frame encoding a protein of 338 residues with a calculated mol. wt of 38,630 and an isoelectric point of 5.85. These results differ from those previously published, which contained a different conceptual amino acid sequence for the carboxy terminal region and identified a different termination codon. Prior to the present invention, the bovine tuftelin protein did not appear to share homology or domain motifs with any other known protein. The gene consists of 13 exons ranging in size from 66 to 1531 bp, the latter containing the encoded carboxyterminal and 3' untranslated regions. The

exons are embedded in than 28 kbp of genomic DNA. Codons generally not divided at exon/intron borders. Several alternatively spliced transcripts were identified by DNA sequence analysis of the isolated products produced by reverse transcriptase/polymerase chain reaction. (See Bashir *et al.*, 1997, Arch Oral Biol 42:489).

- 5 NOV5-7 are new members of the tuftelin-like protein family of proteins. NOV5-7 are useful in detecting fetal liver, testis, fetal lung and thyroid gland tissue. The pattern of expression of NOV5-7 and other tuftelin-like protein family members, and its similarity to the enamelin protein family of genes suggests that it may function as a enamel protein in the tissues of expression. Therefore it is implicated in disorders involving these tissues, such as amelogenesis imperfecta, and other disorders involving enamel defects, including hypoplasia and hypomineralization.
- 10

NOV8

- A NOV8 sequence according to the invention is a nucleic acid sequence encoding a polypeptide related to the neuronal antigen-like protein family of proteins. A NOV8 nucleic acid was mapped to human chromosome 14. A NOV8 nucleic acid is expressed in at least brain, brain stem and testis. A NOV8 nucleic acid and its encoded polypeptide includes the sequences shown in Table 19. The disclosed nucleic acid (SEQ ID NO: 15) is 1,056 nucleotides in length and contains an open reading frame (ORF) that begins with an ATG initiation codon at nucleotides 1-3 and ends with a TGA stop codon at nucleotides 1,054-1,056. The representative ORF encodes a 351 amino acid polypeptide (SEQ ID NO: 16). PSORT analysis predicts that a NOV8 polypeptide is localized to the mitochondrial membrane space (certainty 0.3600)
- 15
- 20

TABLE 19.

- ATGACTTTGAGGCTTTTAGAAGACTGGTGCAGGGGGATGGACATGAACCCTCGGAA
AGCGCTATTGATTGCCGGCATCTCCAGAGCTGCAGTGTGGCAGAAATCGAGGAGG
25 CTCTGCAGGCTGGTTTAGCTCCCTTGGGGGAGTACAGACTGCTTGGAAGGATGTTCA
GGAGGGATGAGAACAGGAAAGTAGCCTTAGTAGGGCTTACTGCGGAGACTAGTCAC
GCCCTGGTCCCTAAGGAGATACCGGGAAGGGGGTATCTGGAGAGTGATCTTTAA
GCCCCCTGACCCAGATAATACATTTTAAAGCAGATTAAATGAATTTTATAGCGGGAGA
GGGCATGACAGTGGGTGAGTTGAGCAGAGCTCTTGGACATGAAAATGGCTCCTTAG
30 ACCCAGAGCAGGGCATGATCCCGGAAATGTGGGCCCTATGTTGGCACAGGCATTA
GAGGCTCTTCAGCCTGCCCTGCAATGCTTGAAGTATAAAAAGCTGAGAGTGTTCTCG
GGCAGGGAGTCTCCAGAACCAGGAGAAGAAGAATTTGGACGCTGGATGTTTCATAC
TACTCAGATGATAAAGGCGTGGCAGGTGCCAGATGTAGAGAAGAGAAGGCGATTGC
TAGAGAGCCTTCGAGGCCAGCACTTGATGTTATTCGTGTCCTCAAGATAAACAATC
35 CTTTAATTACTGTGATGAATGTCTGCAGGCTCTTGAGGAGGTATTTGGGGTTACAG
ATAATCCTAGGGAGTTGCAGGTCAAATATCTAACCCTTACCAGAAGGATGAGGAA
AAGTTGTCGGCTTATGTACTAAGGCTGGAGCCTTTGTTACAGAAGCTGGTACAGAGA

GGAGCAATTGAGACATGCTGTGAATCAGGCCCGCCTACCAAGTCATTGCTGG
GGCAGTCCACAAAACAATTTCGAGAGAGCTTAATCTGCCAGAGGATGGCCCAGCCC
CTGGTTTCTTGCAAGTTATTGGTACTAATAAAGGATTATGAGGCAGCTGAGGAGGAGG
AGGCCCTTCTCCAGGCAATATTGGAAGGTAATTTACCTGA (SEQ ID NO.: 15)

MTLRLLDWCRGMDMNP RKALLIAGISQSCSVAEIEEALQAGLAPLGEYRLLGRMFRR
DENRKVALVGLTAETSHALVPKEIPGKGGIWRVIFKPPDPDNTFLSRLNEFLAGEGMTV
GELSRALGHENGSLDPEQGMIPEMWAPMLAQALEALQPALQCLKYKKLRVFSGRESPE
PGEEEFGRWMFHFTTQMIKAWQVPDVEKRRRLLESLRGPALDVIRVLKINNPLITVDECL
QALEEVFGVTDNPRELQVKYLT TYQKDEEKL SA YVLRLEPLLQKL VQRGAIERDAVNQ
ARLDQVIAGAVHK TIRRELNL PEDGPAPGFLQLL VLIKDYEA AEEEEALLQAILEGNFT
(SEQ ID NO.: 16)

A NOV8 nucleic acid has a high degree of homology (100% identity) with an
uncharacterized region of human chromosome 14 including the clone RPC14-794B2 (CHR14;
GenBank Accession No.: AC005924), as is shown in Table 20. A NOV8 polypeptide has
homology (55% identity, 72% similarity) with a human paraneoplastic neuronal antigen protein
(hPNA; EMBL Accession No.: O95144), as is shown in Table 21. Also, a NOV8 polypeptide has
homology (49% identity, 65% similarity) with a human paraneoplastic neuronal antigen mm2
polypeptide (hPNA; EMBL Accession No.: O95145), as is shown in Table 22. Further, a region
of a NOV8 polypeptide also has a high degree of homology (100% identity) with the human
polypeptide sequence ORF2787 (ORFX; PatP Accession No.: B43023), as shown in Table 39.

TABLE 20.

NOV8 :	1	atgactttgaggcttttagaagactggtgcaggggatggacatgaaccctcggaaagcg	60
HR14 :	19537	atgactttgaggcttttagaagactggtgcaggggatggacatgaaccctcggaaagcg	19478
NOV8 :	61	ctattgattgccggcatctcccagagctgcagtgtggcagaaatcgaggaggctctgcag	120
CHR14 :	19477	ctattgattgccggcatctcccagagctgcagtgtggcagaaatcgaggaggctctgcag	19418
NOV8 :	121	gctggtttagctcccttgggggagtacagactgcttggaggatgttcaggagggatgag	180
CHR14 :	19417	gctggtttagctcccttgggggagtacagactgcttggaggatgttcaggagggatgag	19358
NOV8 :	181	aacaggaaagtagccttagtagggcttactgcggagactagtcacgccctggtccctaag	240
CHR14 :	19357	aacaggaaagtagccttagtagggcttactgcggagactagtcacgccctggtccctaag	19298
NOV8 :	241	gagataccgggaaaaggggtatctggagagtgatctttaagccccctgaccagataat	300
CHR14 :	19297	gagataccgggaaaaggggtatctggagagtgatctttaagccccctgaccagataat	19238
NOV8 :	301	acatttttaagcagattaaatgaatttttagcgggagagggcatgacagtgggtgagttg	360
CHR14 :	19237	acatttttaagcagattaaatgaatttttagcgggagagggcatgacagtgggtgagttg	19178

Chr	Start	End	Seq ID	Seq
NOV8	361	420	1	agcagagggacatgaaaatggctccttagaccagagcagggatccccggaa
CHR14	19177	19118	2	agcagagctcttggacatgaaaatggctccttagaccagagcagggcatgatccccggaa
NOV8	421	480	3	atgtggggccccctatgttggcacaggcattagaggctcttcagcctgccctgcaatgcttg
CHR14	19117	19058	4	atgtggggccccctatgttggcacaggcattagaggctcttcagcctgccctgcaatgcttg
NOV8	481	540	5	aagtataaaaagctgagagtggttctcgggcagggagctctccagaaccaggagaagaagaa
CHR14	19057	18998	6	aagtataaaaagctgagagtggttctcgggcagggagctctccagaaccaggagaagaagaa
NOV8	541	600	7	tttggacgctggatgtttcatactactcagatgataaaggcggtggcaggtgccagatgta
CHR14	18997	18938	8	tttggacgctggatgtttcatactactcagatgataaaggcggtggcaggtgccagatgta
NOV8	601	660	9	gagaagagaaggcgattgctagagagccttcgaggccagcacttgatgttattcgtgtc
CHR14	18937	18878	10	gagaagagaaggcgattgctagagagccttcgaggccagcacttgatgttattcgtgtc
NOV8	661	720	11	ctcaagataaacaatcctttaattactgtcgatgaatgtctgcaggctcttgaggaggtta
CHR14	18877	18818	12	ctcaagataaacaatcctttaattactgtcgatgaatgtctgcaggctcttgaggaggtta
NOV8	721	780	13	tttgggggttacagataatcctagggagttgcaggtcaaatatctaaccacttaccagaag
CHR14	18817	18758	14	tttgggggttacagataatcctagggagttgcaggtcaaatatctaaccacttaccagaag
NOV8	781	840	15	gatgaggaaaagtgtgcggccttatgtactaaggctggagcctttgttacagaagctggta
CHR14	18757	18698	16	gatgaggaaaagtgtgcggccttatgtactaaggctggagcctttgttacagaagctggta
NOV8	841	900	17	cagagaggagcaattgagagagatgctgtgaatcaggcccgcttagaccaagtcatgtct
CHR14	18697	18638	18	cagagaggagcaattgagagagatgctgtgaatcaggcccgcttagaccaagtcatgtct
NOV8	901	960	19	ggggcagtcacacaaaacaattcgagagagcttaatctgccagaggatggcccagcccct
CHR14	18637	18578	20	ggggcagtcacacaaaacaattcgagagagcttaatctgccagaggatggcccagcccct
NOV8	961	1020	21	ggtttcttgacgttattgggtactaataaaggattatgaggcagctgaggaggaggaggcc
CHR14	18577	18518	22	ggtttcttgacgttattgggtactaataaaggattatgaggcagctgaggaggaggaggcc
NOV8	1021	1056 (SEQ ID NO.:15)	23	cttctccaggcaatattggaagtaatttcacctga
CHR14	18517	18482 (SEQ ID NO.:38)	24	cttctccaggcaatattggaagtaatttcacctga

TABLE 21.

```

NOV8:      1  MTLRLLEDWCRGMDMNP RKALLIAGISQSCSVAEIEEALQAGLAPLGEYRL LGRMFRDE  60
          * + *****++ ++ **+ ** +* ***** ** + * **+***** **
HPNA:      1  MAMTLLEDWCRGMDVNSQRXLLVWGIPVNCDEAEIEETLQAAM-PQVSYRMLGRMFWRREE  59

NOV8:     61  NRKVALVGLTAETSHALVPKEIPGKGGIWRVIFKPPDPDN TFLSRLNEFLAGEGMTV GEL  120
          * * **+ ** +* ++*+*****++*+***** * ** **+ ** ** ** ++
HPNA:     60  NAKAALLELTGAVDYAAIPREMPGKGGVWKVLFKPP TSDAEFLERLHLFLAREGWTVDV  119

```

NOV8: 121 SRALGHELDPEQGMIPEMWAPMLAQALE-ALQPALQCLKRVFSGRESPEPGEE 179
 +* ** +* + * * ** * ** *+ *** ++ + **++ +**** *

HPNA: 120 ARVLGFQNPPT--PTPG--PEMPAEMLNLYILDNVIQPLVESIWKRLTLFSGKGHPRAWRG 175

5 NOV8: 180 EFGRWMFHTTQMIKAWQVPDVEKRRRLLESRLGPALDVIRVLKINNPLITVDECLQALEE 239
 * **++ ** +++++ ** * ****+***** ****+** ** * ****+****

HPNA: 176 NFDPWLEHTNEVLEEWQVSDVEKRRRLMESLRGPAADVIRILKSNNPAITTAECKALEQ 235

10 NOV8: 240 VFGVTDNPRELQVKYLTQKDEEKL SAYVLRLEPLLQKLVQRGAIERDAVNQARLDQVI 299
 *** ++ ** +++++ ** * ****+***** ****+** ** * ****+****

HPNA: 236 VFGSVESRDAQIKFLNTYQNPGEKLSAYVIRLEPLLQKVVEKGAIDKDNVNQARLEQVI 295

NOV8: 300 AGAVHK-TIRRELNLPEDGPAPG 321 (SEQ ID NO.:39)
 *** * ****+* * * **

15 HPNA: 296 AGANHSGAIRRLWLTGAGEGPG 318 (SEQ ID NO.:40)
 Where * indicates identity and + indicates similarity.

TABLE 22.

20 NOV8 : 3 LRLLEDWCRGMDMNP RKALLIAGISQSCSVAEIEEALQAGLAPLGEYRLLGRMFRRDENR 62
 * **** * **++ +++++ ** * ****+***** ****+** ** * ****+****

HPNA: 1 LALLEDWCRIMSVDEQKSLMVTGIPADFBEEAIQEVLTSLGRYRLLGKIFRKQENA 60

NOV8 : 63 KVALVGLTAETSHALVPKEIPGKGGIWRVIFKPPDPDNTFLSRLNEFLAGEGMTVGELSR 122
 (SEQ ID NO.: 41)
 *+ * ** +* **++ ****+***** ** * ** ** ** ** ** ** ** * ** * + *

25 HPNA: 61 NAVLLELLEDTDVSAIPSEVQKGGVWVKVIFKTPNQDTEFLERLNLFLKEGQTVSGMFR 120
 (SEQ ID NO.: 42)
 Where * indicates identity and + indicates similarity.

TABLE 39.

30 NOV8: 1 MTLRLLEDWCRGMDMNP RKALLIAGISQSCSVAEIEEALQAGLAPLGEYRLLGRMFRRDE 60

ORFX: 1 MTLRLLEDWCRGMDMNP RKALLIAGISQSCSVAEIEEALQAGLAPLGEYRLLGRMFRRDE 60

NOV8: 181 NRKVALVGLTAETSHALVPKEIPGKGGIWRVIFKPPDPDNTFLSRLNEFLAGEGMTVGEL 120

35 ORFX: 61 NRKVALVGLTAETSHALVPKEIPGKGGIWRVIFKPPDPDNTFLSRLNEFLAGEGMTVGEL 120

NOV8: 361 SRALGHENGLDPEQGMIPEMWAPMLAQALEALQPALQCLKYKLRVFSGRESPEPGEEE 180

40 ORFX: 121 SRALGHENGLDPEQGMIPEMWAPMLAQALEALQPALQCLKYKLRVFSGRESPEPGEEE 180

NOV8: 541 FGRWMFHTTQMIKAWQVPDVEKRRRLLESRLGPALDVIRVLKINNPLITVDECLQALEEV 240

ORFX: 181 FGRWMFHTTQMIKAWQVPDVEKRRRLLESRLGPALDVIRVLKINNPLITVDECLQALEEV 240

45 NOV8: 721 FGVTDNPRELQVKYLTQKDEEKL SAYVLRLEPLLQKLVQRGAIERDAVNQARLDQVIA 300

ORFX: 241 FGVTDNPRELQVKYLTQKDEEKL SAYVLRLEPLLQKLVQRGAIERDAVNQARLDQVIA 300

NOV8: 901 GAVHKTIRRELN 312 (SEQ ID NO.: 73)

ORFX: 301 GAVHKTIRRELN 312 (SEQ ID NO.: 74)

Where * indicates identity.

5 In patients with cancer, symptoms of limbic and brain-stem dysfunction may result from a paraneoplastic disorder. Paraneoplastic limbic or brain-stem encephalitis occurs more frequently with testicular cancer than with most other cancers. Antineuronal antibodies may be used in a diagnostic test for this syndrome. Immunohistochemical and immunoblotting techniques were used to detect serum and cerebrospinal fluid antibodies. Serologic screening of a complementary DNA library and Northern blotting have been used to clone the target antigen and determine which tissues expressed it. Of 13 patients with testicular cancer and paraneoplastic limbic or brain-stem encephalitis (or both), 10 had antibodies in serum and cerebrospinal fluid against a 40-kd neuronal protein. These antibodies were used to clone a gene called Ma2, which codes for a protein (Ma2) that was recognized by serum from the 10 patients, but not by serum from 344 control subjects. Ma2 was selectively expressed by normal brain tissue and by the testicular tumors of the patients. Ma2 shares homology with Ma1, a "brain-testis-cancer" gene related to other paraneoplastic syndromes and tumors. Therefore, the serum of patients with subacute limbic and brain-stem dysfunction and testicular cancer contains antibodies against a protein found in normal brain and in testicular tumors. Detection of these antibodies supports the paraneoplastic origin of the neurologic disorder and could be of diagnostic importance. (See Voltz *et al.*, 1999, N. Engl. J. Med. 340:1788.)

Also, the identification of antineuronal antibodies has facilitated the diagnosis of paraneoplastic neurological disorders and the early detection of the associated tumours. It has also led to the cloning of possibly important neuron-specific proteins. Serological studies of 1705 sera from patients with suspected paraneoplastic neurological disorders resulted in the identification of four patients with antibodies that reacted with 37 and 40 kDa neuronal proteins (anti-Ma antibodies). Three patients had brainstem and cerebellar dysfunction, and one had dysphagia and motor weakness. Autopsy of two patients showed loss of Purkinje cells, Bergmann gliosis and deep cerebellar white matter inflammatory infiltrates. Extensive neuronal degeneration, gliosis and infiltrates mainly composed of CD8+ T cells were also found in the brainstem of one patient. In normal human and rat tissues, the anti-Ma antibodies reacted exclusively with neurons and with testicular germ cells; the reaction was mainly with subnuclear elements (including the nucleoli) and to a lesser degree the cytoplasm. Anti-Ma antibodies also

reacted with the cancers (breast, colon and parotid) available from the anti-Ma patients, but not with 66 other tumours of varying histological types. Preincubation of tissues with any of the anti-Ma sera abrogated the reactivity of the other anti-Ma immunoglobulins. Probing of a human complementary DNA library with anti-Ma serum resulted in the cloning of a gene that encodes a novel 37 kDa protein (Mal). Recombinant Mal was specifically recognized by the four anti-Ma sera but not by 337 control sera, including those from 52 normal individuals, 179 cancer patients without paraneoplastic neurological symptoms, 96 patients with paraneoplastic syndromes and 10 patients with non-cancer-related neurological disorders. The expression of Mal mRNA is highly restricted to the brain and testis. Subsequent analysis suggested that Mal is likely to be a phosphoprotein. Some patients with paraneoplastic neurological disorders develop antibodies against Mal, a new member of an expanding family of 'brain/testis' proteins.

NOV8 is a new member of the neuronal antigen-like protein family of proteins. The pattern of expression of the NOV8 gene and its family members, and its similarity to the neuronal antigen-like protein family of genes suggests that it may function as a neuronal antigen in the tissues of expression. NOV8 is useful as a marker for brain, brainstem and testis tissue. Therefore NOV8 is implicated in disorders involving these tissues. Some of the diseases include but are not limited to: cardiovascular disorders, diabetes, leukemia/lymphoma, cancer, musculoskeletal disorders, muscular degeneration, reproductive health, metabolic and endocrine disorders, gastrointestinal disorders, immune disorders and Autoimmune diseases, respiratory disorders, bone disorders, and tissue/Cell growth regulation disorders. NOV8 is also useful as a marker for human chromosome 14.

NOV9

A NOV9 sequence according to the invention is a nucleic acid sequence encoding a polypeptide related to the fatty acid binding protein-like protein family of proteins. A NOV9 nucleic acid and its encoded polypeptide includes the sequences shown in Table 23. The disclosed nucleic acid (SEQ ID NO: 17) is 499 nucleotides in length and contains an open reading frame (ORF) that begins with an ATG initiation codon at nucleotides 5-7 and ends with a TAA stop codon at nucleotides 494-496. The representative ORF encodes a 163 amino acid polypeptide (SEQ ID NO: 18). Putative untranslated regions upstream and downstream of the coding sequence are underlined in SEQ ID NO: 17. PSORT analysis predicts that a NOV9

polypeptide is a cytoplasmic protein (certainty 0.6500). SIGNALP analysis suggests that a NOV9 polypeptide lacks a signal peptide.

TABLE 23.

CAAAATGGTTAAGAACACAAACCAGTACGCTGCTCACGCCGATCCCGCTCCGCTGG
TTCCGCACGCTCCGCACACCAGCCTGCGCGCACCATGGGCCACCGTTCAGCAGCTGG
AAGGAAGATGGCGCCTGGCGGACAGCAAAGGCTTTGATGCATACATGAAGAACTA
GGAGTGGGAATATCTTTGCGCAATATGGGCGCAATGGCCAAACCAGACTGTATCAT
CACTTGTGATGGCAAAAACCTCACCATAAAAACTGAGAGCACTTTGAAAACAACAC
AGTTTTCTTGTACCCTGGGAGAGAAGTTTGAAGGAACCACAGCTGTTGGCAGAAAA
ACTCAGACTGTCTGCAGCTTTACAGATGGTGCATTGGTTCCGCATCAGGAGTGGGAT
GGGAAGGAAAACACAATAACAAGAAAATTGAAAGATGCATCAGTGGTGGATTGTGT
CACGAACAATGTCACCTGTACTCGGATCTATGAAAAAGTAGAATAAAAAA (SEQ ID
NO.: 17)

MVKNTNQYAAHADPAPLVPHAPHTSLRAPWATVQQLEGRWRLADSKGFDAYMKKLG
VGISLRNMGAMAKPDCIITCDGKNLTIKTESTLKTTFQFCTLGEKFEGTTAVGRKTQTV
SFTDGLVPHOEWDGKENTITRKLKDasVDCVTNNVTCTRIYEKVE (SEQ ID NO.: 18)

TABLE 24.

A NOV9 nucleic acid has a high degree of homology (92% identity) with a human fatty acid binding protein homolog mRNA (hFBP; GenBank Accession No.: M94856), as is shown in table 24. A NOV9 nucleic acid also has a high degree of homology (94% identity) with a human melanogenic inhibitor mRNA (hMI; PatP Accession No.: R55866), as is shown in Table 25. A NOV9 polypeptide has homology (88% identity, 92% similarity) with a human epidermal fatty acid-binding protein (eFBP; SwissProt Accession No.: Q01469), as is shown in Table 26.

NOV9: 359 ACAGATGGT TGGTTCCGCATCAGGAGTGGGATGGGAAGGAAAAC TAACAAGA 418
 NFBP: 316 ACAGATGGTGCATTGGTTCCAGCATCAGGAGTGGGATGGGAAGGAAAGCACATAACAAGA 375

5 NOV9: 419 AAATTGAAAGATGC-A--TCAGTGGTGGATTGTGTACGAACAATGTACCTGTACTCGG 475
 NFBP: 376 AAATTGAAAGATGGGAAATTAGTGGTGGAGTGTGTATGAACAATGTACCTGTACTCGG 435

10 NOV9: 476 ATCTATGAAAAAGTAGAATAAAAA 499 (SEQ ID NO.: 43)
 NFBP: 436 ATCTATGAAAAAGTAGAATAAAAA 459 (SEQ ID NO.: 44)

TABLE 25.

NOV9: 94 GGCCACCGTTCAGCAGCTGGAAGGAAGATGGCGCCTGGCGGACAGCAAAGGCTTTGATGC 153
 HMI: 3 GGCCACAGTTCAGCAGCTGGAAGGAAGATGGCGCCTGGTGGACAGCAAAGGCTTTGATGA 62

NOV9: 154 ATACATGAAGAACTAGGAGTGGGAATATCTTTGCGCAATATGGGCGCAATGGCCAAACC 213
 HMI: 63 ATACATGAAGGAGCTAGGAGTGGGAATAGCTTTGCGAAAAATGGGCGCAATGGCCAAGCC 122

NOV9: 214 AGACTGTATCATCACTTGTGATGGCAAAACCTCACCATAAAAACTGAGAGCACTTTGAA 273
 HMI: 123 AGATTGTATCATCACTTGTGATGGTAAAAACCTCACCATAAAAACTGAGAGCACTTTGAA 182

NOV9: 274 AACAAACACAGTTTTCTGTACCTGGGAGAGAAGTTGAAGGAACCACAGCTGTGGCAG 333
 HMI: 183 AACAAACACAGTTTTCTGTACCTGGGAGAGAAGTTGAAGAAACCACAGCTGTGGCAG 242

NOV9: 334 AAAAAGTCTGAGTGTCTGAGCTTTACAGATGGTGCATTGGTTCCGCATCAGGAGTGGGA 393
 HMI: 243 AAAAAGTCTGAGTGTCTGCACTTTACAGATGGTGCATTGGTTCAGCATCAGGAGTGGGA 302

NOV9: 394 TGGGAAGGAAACACAATAACAAGAAAATTGAAAGATGC-A--TCAGTGGTGGATTGTGT 450
 HMI: 303 TGGGAAGGAAAGCACATAACAAGAAAATTGAAAGATGGGAAATTAGTGGTGGAGTGTGT 362

NOV9: 451 CACGAACAATGTACCTGTACTCGGATCTATGAAAAAGTAGAATAA 496 (SEQ ID NO.: 45)
 HMI: 363 CATGAACAATGTACCTGTACTCGGATCTATGAAAAAGTAGAATAA 408 (SEQ ID NO.: 46)

TABLE 26.

NOV9: 31 ATVQQLGRWRLADSKGFDAYMKLGVGISLRNMGAMAKPDCIITCDGKNLTIKTESTLK 90
 *****+*****
 Sbjet: 2 ATVQQLGRWRLVDSKGFDEYMKELGVGIALRKMAMAKPDCIITCDGKNLTIKTESTLK 61

NOV9: 91 TTQFSTLGEKFEGTTAVGRKTQTVCSFTD GALVPHQEWGKENTITRKLK DAS-VVDCV 150
 *****+*****+*****
 Sbjet: 62 TTQFSTLGEKFEETTADGRKTQTVCNFTD GALVQHQEWGKESTITRKLKDGKLVVBCV 121

NOV9: 151 TNNVTCTRIYEKVE 164 (SEQ ID NO.: 47)

 Sbjet: 122 MNNVTCTRIYEKVE 135 (SEQ ID NO.: 48)

Where * indicates identity and + indicates similarity.

NOV10

A NOV10 sequence according to the invention is a nucleic acid sequence encoding a polypeptide related to the fatty acid binding protein-like protein family of proteins. A NOV10 nucleic acid and its encoded polypeptide includes the sequences shown in Table 27. The disclosed nucleic acid (SEQ ID NO: 19) is 413 nucleotides in length and contains an open reading frame (ORF) that begins with an ATG initiation codon at nucleotides 6-8 and ends with a TAA stop codon at nucleotides 408-410. The representative ORF encodes a 134 amino acid polypeptide (SEQ ID NO: 20). PSORT analysis suggests that a NOV10 polypeptide localizes to the cytoplasm (certainty 0.6500) and SIGNALP analysis suggests that the NOV10 polypeptide has no signal peptide. Putative untranslated regions upstream and downstream of the coding sequence are underlined in SEQ ID NO: 19.

TABLE 27.

GCACCATGGCCACCGTTCAGCAGCTGGAAGGAAGATGGCGCCTGGCGGACAGCAAA
GGCTTTGATGCATACATGAAGAACTAGGAGTGGGAATATCTTTGCGCAATATGGGC
15 GCAATGGCCAAACCAGACTGTATCATCACTTGTGATGGCAAAAACCTCACCATAAA
AACTGAGAGCACTTTGAAAACAACACAGTTTTCTTGTACCCTGGGAGAGAAGTTTGA
AGGAACCACAGCTGTTGGCAGAAAACTCAGACTGTCTGCAGCTTTACAGATGGTG
CATTGGTTCCGCATCAGGAGTGGGATGGGAAGGAAAACACAATAACAAGAAAATTG
20 AAAGATGCATCAGTGGTGGATTGTGTACGAACAATGTCACCTGTACTCGGATCTAT
GAAAAAGTAGAATAAAAA (SEQ ID NO.: 19)

MATVQQLEGRWRLADSKGFDAYMKKLGVGISLRNMGAMAKPDCIITCDGKNLTIKTES
TLKTTQFSTLGEKFEGTTAVGRKTQTVCSFTDGALVPHQEWDGKENTITRKLKDASVV
DCVTNNVTCTRIYEKVE (SEQ ID NO.: 20)

A NOV10 nucleic acid has a high degree of homology (94% identity) with a human fatty acid binding protein homolog mRNA (GenBank Accession No.: M94856). A NOV10 nucleic acid also has a high degree of homology (94% identity) with a human melanogenic inhibitor mRNA (PatP Accession No.: R55866). A NOV10 polypeptide has homology (88% identity, 92% similarity) with a human epidermal fatty acid-binding protein (eFBP; SwissProt Accession No.: Q01469), as is shown in Table 28.

TABLE 28.

NOV10: 1 MATVQQLEGRWRLADSKGFDAYMKKLGVGISLRNMGAMAKPDCIITCDGKNLTIKTESTL 60

35 EFBP: 1 MATVQQLEGRWRLVDSKGFDAYMKELGVGIALRKMAMAKPDCIITCDGKNLTIKTESTL 60
NOV10: 61 KTTQFSTLGEKFEGTTAVGRKTQTVCSFTDGALVPHQEWDGKENTITRKLKDAS-VVDC 120

40 EFBP: 61 KTTQFSTLGEKFEEETADGRKTQTVCSFTDGALVPHQEWDGKESTITRKLKDGKLVVEC 120

NOV10: 121 VTNNVTCTRIYEKVE 135 (SEQ ID NO.: 20)
 * *****
 EFBP: 121 VMNNVTCTRIYEKVE 135 (SEQ ID NO.: 49)

5 Where * indicates identity and + indicates similarity.

NOV9-10 are highly homologous to each other and to other members of the fatty acid-binding protein-like family of proteins, as is shown by CLUSTALW analysis in Table 29.

TABLE 29.

NOV10 -----MATVQQLEGRWRLADSKGFDAYMKKLGVGIS
 NOV9 MVKNTNQYAAHADPAPLVPHPHTSLRAPWATVQQLEGRWRLADSKGFDAYMKKLGVGIS
 hFABP -----MATVQQLEGRWRLVDSKGFDEYMKELGVGIA
 rFABP -----MASLKDLEGKWRLVESHGFEYMKELGVGLA
 mFABP -----MASLKDLEGKWRLMESHGFEEYMKELGVGLA
 *:::***:*** :*:***: ***:*****:

NOV10 LRNMGAMAKPDCIITCDGKNLTIKTESTLKTTFQFCTLGEKFEGTTAVGRKTQTVCSFTD
 NOV9 LRNMGAMAKPDCIITCDGKNLTIKTESTLKTTFQFCTLGEKFEGTTAVGRKTQTVCSFTD
 hFABP LRKMGAMAKPDCIITCDGKNLTIKTESTLKTTFQFCTLGEKFEBTTADGRKTQTVCNFTD
 rFABP LRKMGAMAKPDCIITLDGNNLTIVKTESTVKTTFVFSCTLGEKFDETTADGRKTETVCTFTD
 mFABP LRKMAAMAKPDCIITCDGNNITVKTESTVKTTFVFSCTLGEKFDETTADGRKTETVCTFQD
 :*.....:*:*****:*** ***.*****: *** *****:***.* *

NOV10 GALVPHQEWGDKENTITRKLKDAS-VVDCVTNNVTCTRIYEKVE (SEQ ID NO.: 20)
 NOV9 GALVPHQEWGDKENTITRKLKDAS-VVDCVTNNVTCTRIYEKVE (SEQ ID NO.: 18)
 hFABP GALVQHQEWGDKESTITRKLKDGKLVVECVMMNNVTCTRIYEKVE (SEQ ID NO.: 50)
 rFABP GALVQHQEWGDKESTITRKLKDGKLVVECVMMNAICTRVYEKVV (SEQ ID NO.: 51)
 mFABP GALVQHQQWDGKESTITRKLKDGKMIVECVMMNATCTRVYEKVV (SEQ ID NO.: 52)
 **** *:*:***.*****. . :*:** *. *****:

Where * indicates identity, : indicates strong similarity and . indicates weak homology.

Rat fatty acid-binding protein (rFABP; SwissProt Accession No.: P55053), mouse fatty acid-binding protein (mFABP; SwissProt Accession No.: Q05816), and human fatty acid-binding protein (hFABP; SwissProt Accession No.: Q01469).

Fatty acid metabolism in mammalian cells depends on a flux of fatty acids, between the plasma membrane and mitochondria or peroxisomes for beta-oxidation, and between other cellular organelles for lipid synthesis. The fatty acid-binding protein (FABP) family consists of small, cytosolic proteins believed to be involved in the uptake, transport, and solubilization of their hydrophobic ligands. Members of this family have highly conserved sequences and tertiary structures. Fatty acid-binding proteins were first isolated in the intestine (FABP2; OMIM-134640) and later found in liver (FABP1; OMIM-134650), striated muscle (FABP3; OMIM-

134651), adipocytes (FABP4; OMIM- 600434) and epidermal tissues (E-FABP; GDB ID:136450).

Epidermal fatty acid binding protein (E-FABP) was cloned by as a novel keratinocyte protein by Madsen and co-workers (See Madsen *et al.*, 1992, *J. Invest. Dermatol.* 99:299) from skin of psoriasis patients. Later, using quantitative Western blot analysis Kingma *et al.* (See Kingma *et al.*, 1998, *Biochemistry* 37:3250) have shown that in addition to the skin, bovine E-FABP is expressed in retina, testis, and lens. Since E-FABP was originally identified from the skin of psoriasis patients, it is also known as psoriasis-associated fatty acid-binding protein (PA-FABP). PA-FABP is a cytoplasmic protein, and is expressed in keratinocytes. It is highly up-regulated in psoriatic skin. It shares similarity to other members of the fatty acid-binding proteins and belongs to the fabp/p2/crbp/crabp family of transporter. PA-FABP is believed to have a high specificity for fatty acids, with highest affinity for c18 chain length. Decreasing the chain length or introducing double bonds reduces the affinity. PA-FABP may be involved in keratinocyte differentiation.

Immunohistochemical localization of the expression of E-FABP in psoriasis, basal and squamous cell carcinomas has been carried out in order to obtain indirect information, at the cellular level, on the transport of the fatty acids. (See Masouye *et al.*, 1996, *Dermatology* 192:208). E-FABP was localized in the upper stratum spinosum and stratum granulosum in normal and non-lesional psoriatic skin. In contrast, lesional psoriatic epidermis strongly expressed E-FABP in all suprabasal layers, like nonkeratinized oral mucosa. The basal layer did not express E-FABP reactivity in any of these samples. Accordingly, basal cell carcinomas were E-FABP negative whereas only well-differentiated cells of squamous cell carcinomas expressed E-FABP. This suggests that E-FABP expression is related to the commitment of keratinocyte differentiation and that the putative role of E-FABP should not be restricted to the formation of the skin lipid barrier. Since the pattern of E-FABP expression mimics cellular FA transport, lesional psoriatic skin and oral mucosa may have a higher metabolism/transport for FAs than normal and non-lesional psoriatic epidermis.

NOV9-10 represent new members of a family of epidermal fatty acid-binding proteins, and are thus useful to determine epidermal fatty acid-binding protein interacting proteins. The pattern of expression of the NOV9-10 genes and their family members, and their similarity to the epidermal fatty acid-binding protein family of genes suggests that it may function as a regulator

of fatty acid synthesis, uptake, transport, localization and solubilization of their hydrophobic ligands in the tissues of expression, *e.g.* keratinocytes. NOV9-10 are useful as markers for keratinocyte differentiation. Therefore NOV9-10 are implicated in disorders involving these tissues. Some of the diseases include but are not limited to: lesional psoriatic skin and oral mucosa.

NOV11

A NOV11 sequence according to the invention is a nucleic acid sequence encoding a polypeptide related to the cystatin family of proteins. A NOV11 nucleic acid and its encoded polypeptide includes the sequences shown in Table 30. The disclosed nucleic acid (SEQ ID NO: 21) is 468 nucleotides in length and contains an open reading frame (ORF) that begins with an ATG initiation codon at nucleotides 10-12 and ends with a TGA stop codon at nucleotides 445-447. The representative ORF encodes a 145 amino acid polypeptide (SEQ ID NO: 22). A mature form of a NOV11 polypeptide is described in Example 3. Putative untranslated regions upstream and downstream of the coding sequence are underlined in SEQ ID NO: 21. PSORT analysis predicts that a NOV11 polypeptide is a secreted protein (certainty 0.6042), and SIGNALP analysis suggests that NOV11 has an N-terminal signal peptide, most likely between positions 19 and 20 of SEQ ID NO.: 22.

TABLE 30.

GCTGTAGACATGGGGATCGGATGCTGGAGAAACCCCTGCTGCTGCTGATTGCCCTG
GTCCTGTCAGCCAAGCTGGGTCACTTCCAAAGGTGGGAGGGCTTCCAGCAGAAGCT
CATGAGCAAGAAGAACATGAATTCAACACTCAACTTCTTCATTCAATCCTACAACAA
TGCCAGCAACGACACCTACTTATATCGAGTCCAGAGGCTAATTCGAAGTCAGATGCA
GCTGACGACGGGAGTGGAGTATATAGTCACTGTGAAGATTGGCTGGACCAAATGCA
AGAGGAATGACACGAGCAATTCTTCTGCCCCCTGCAAACCAAGAAGCTGAGAAAG
AGTTTAATTTGCGAGTCTTTAATATACACCATGCCCTGGTTAAACTATTTCAGCTCT
GGAACAATTCCTGTCTGGAGCCCGAGCATGTGGGCAGAAACCTCAGATGAGGGCTC
ATATGATTGAGTTGT (SEQ ID NO.: 21)

MGIGCWRNPLLLLIALLVLSAKLGHFQRWEGFQQKLMSKKNMNSTLNFFIQSYNNASND
TYLYRVQRLIRSQMQLTTGVEYIVTVKIGWTKCKRNDTSNSSCPLQTKLRKSLICESLI
YTMPWLNYFQLWNNSCLEPEHVGRLNR (SEQ ID NO.: 22)

A NOV11 nucleic acid has a high degree of homology (100% identity) with a region of human chromosome 20p11.21-12.3, including the clone RP3-322G13 (CHR20; GenBank Accession No.: HSJ322G13), as is shown in Table 31. A NOV11 polypeptide is homologous to a rat cystatin C polypeptide (RCYS; GenBank Accession No.: P14841), as is shown in Table 32.

NOV11 is also homologous to a human cystatin D polypeptide (hCYD; GenBank Accession No.: P28325) as is shown in Table 33.

TABLE 31.

5	NOV11: 1	gctgtagacatggggatcggatgctggagaaacccctgctgctgctgattgccctggtc	60
	CHR20: 95473	gctgtagacatggggatcggatgctggagaaacccctgctgctgctgattgccctggtc	95532
	NOV11: 61	ctgtcagccaagctgggtcacttccaaagtgagggtccagcagaagctcatgagc	120
10	CHR20: 95533	ctgtcagccaagctgggtcacttccaaagtgagggtccagcagaagctcatgagc	95592
	NOV11: 121	aagaagaacatgaattcaacactcaacttcttcattcaatcctacaacaatgccagcaac	180
15	CHR20: 95593	aagaagaacatgaattcaacactcaacttcttcattcaatcctacaacaatgccagcaac	95652
	NOV11: 181	gacacctacttatatcgagtcagaggctaattcgaagtcagatgcag	228 (SEQ ID NO.: 53)
	CHR20: 95653	gacacctacttatatcgagtcagaggctaattcgaagtcagatgcag	95700 (SEQ ID NO.: 54)

TABLE 32.

20	NOV11: 37	SKKNMNSTLNFFIQSYNNASNDTYLYRVQRLIRSQMQLTTGVEYIVTVKIGWTKCKRNDT	97
	RCYS: 19	SEEGVQRALDFAVSEYNKGSNDAYHSRAIQVVRARKQLVAGINYLDVEMGRITCTKSQT	78
25	NOV11: 98	SNSSCPL--QTKLRKSLICESLIYTMPWLNYPQLWNNSC	134 (SEQ ID NO.: 55)
	RCYS: 79	NLTNCPFHDQPHLMRKAL-CSFQIYVPWKGTHTLTKSSC	117 (SEQ ID NO.: 56)

Where * indicates identity and + indicates similarity.

TABLE 33.

30	NOV11: 36	MSKKNMNSTLNFFIQSYNNASN-DTYLYRVQRLIRSQMQLTTGVEYIVTVKIGWTKCKRN	94
	HCYS: 39	LNDKSVQCALDFAISEYNKVINKDEYYSRPLQVMAAYQQIVGGVNYFNVKFGRTTCTKS	98
35	NOV11: 95	DTSNSSCPLQTK-KLRKSLICESLIYTMPWLNYPQLWNNSC	134 (SEQ ID NO.: 57)
	HCYS: 99	QPNLDNCPFNDQPKLKEEEFCFSFQINEVPWEDKISILNYKC	139 (SEQ ID NO.: 58)

Where * indicates identity and + indicates similarity.

The superfamily of cysteine-proteinase inhibitors comprises structurally homologous proteins which are divided into at least three families: family I (the stefins; see OMIM 184600), family II (the cystatins), and family III (the kininogens; see OMIM 228960). Salivary cystatins, known as cystatins S, SA and SN, are mainly found in saliva, tears, and seminal plasma, whereas cystatin C (CST3; OMIM 604312) is abundant in cerebrospinal fluid, seminal plasma, milk, synovial fluid, and urine and blood plasma of patients with uremia. It is possible that these proteins play important roles in the protection of cells from inappropriate proteolysis and in the regulation of cysteine-proteinases of both host and bacterial origin. Saitoh and co-workers (See

Saitoh, et al., 1987, Genes 1:329) showed that the salivary-type cystatins are determined by a gene family that consists of at least seven loci. They isolated three cystatin genes, CST1 for cystatin SN, CST2 for cystatin SA, and CSTEP1 for a cystatin pseudogene. Saitoh *et al.* demonstrated that CST3, the gene that codes for cystatin C and is mutant in cerebral amyloid angiopathy of the Icelandic type (OMIM 105150), has the same organization as the CST1 and CST2 genes (See Saitoh *et al.*, 1989, Biochem. Biophys. Res. Commun. 162:1324). Southern analysis of somatic cell hybrid clones demonstrated that all members of the cystatin gene family segregate with human chromosome 20.

Cystatin C, which belongs to the type II cystatin gene family, is the most abundant extracellular inhibitor of cysteine proteases. It is a 13-kD protein constitutively secreted shortly after its synthesis (See Barrett *et al.*, 1984, Biochem. Biophys. Res. Commun. 120:631). Grubb and Lofberg (See Grubb and Lofberg, 1982, Proc. Nat. Acad. Sci. 79:3024) reported the amino acid sequence of the protein. The isolation and characterization of six human cysteine proteinase inhibitors, including cystatin C was reported in 1988 (See Abrahamson, 1988, Scand. J. Clin. Lab. Invest. Suppl. 191:21). Whereas cystatins D (OMIM 123858), S (OMIM 123857), and SA (OMIM 123856) are expressed primarily in salivary glands, cystatin C is expressed in virtually all organs of the body. According to its high concentration in biologic fluids, cystatin C is probably one of the most important extracellular inhibitors of cysteine proteases. Cystatin C is present in a number of neuroendocrine cells and its concentration in the cerebrospinal fluid is 5.5 times that in plasma of healthy adults (See Grubb and Lofberg, 1982, Proc. Nat. Acad. Sci. 79:3024).

The pathogenesis of atherosclerosis and abdominal aortic aneurysm (AAA; OMIM 100070) involves breakdown of the elastic laminae. Elastolytic cysteine proteases, including cathepsins S and K, are overexpressed at sites of arterial elastin damage, but whether endogenous local inhibitors counterbalance these proteases was unknown. Shi and colleagues (See Shi *et al.*, 1999, J. Clin. Invest. 104:1191) showed that, whereas cystatin C is normally expressed in vascular wall smooth muscle cells, this cysteine protease inhibitor is severely reduced in both atherosclerotic and aneurysmal aortic lesions. Furthermore, increased abdominal aortic diameter among 122 patients screened by ultrasonography correlated inversely with serum cystatin C levels. In vitro, cytokine-stimulated vascular smooth muscle cells secrete cathepsins, whose elastolytic activity could be blocked when cystatin C secretion was induced by treatment with

TGF-beta-1. These findings highlighted a potentially important role in imbalance between cysteine proteases and cystatin C in arterial wall remodeling and established that cystatin C deficiency occurs in vascular disease. Shi *et al.* also reported that, to their knowledge, the marked suppression of cystatin C concurrent with augmented expression of cysteine proteases observed in their studies of atherosclerosis and abdominal aneurysms represented the first acquired cysteine protease inhibitor deficiency in human disease (See Shi *et al.*, 1999, J. Clin. Invest. 104:1191).

NOV11 represents a new member of the cystatin family of proteins. As such, NOV11 is useful in identifying cystatin-interacting proteins. NOV11 is also useful as a marker for the region of human chromosome 20p11.21-12.3.

The nucleic acids and proteins of the invention are useful in potential therapeutic applications implicated in disorders characterized by aberrant cell proliferation, differentiation and migration, *e.g.* cancer, angiogenesis and wound healing, neurological disorders, *e.g.* paraneoplastic neurological disorders, episodic ataxia, autosomal dominant myokymia, stroke, Parkinson's disease and Alzheimer's disease, enamel defects, *e.g.* amelogenesis imperfecta, and inappropriate proteolysis, *e.g.* atherosclerosis and abdominal aortic aneurysms. For example, a cDNA encoding a tuftelin-like protein may be useful in gene therapy for treating amelogenesis imperfecta and other such disorders, and the tuftelin-like protein may be useful when administered to a subject in need thereof. By way of nonlimiting example, the compositions of the present invention will have efficacy for treatment of patients suffering from disorders of the neurological system. The novel nucleic acids encoding a potassium channel-like protein, and the potassium channel-like protein of the invention, or fragments thereof, may further be useful in the treatment of Episodic Ataxia, type 1, Long QT Syndrome 1 and 2, Benign Neonatal Epilepsy, Jervell and Lange-Neilson syndrome, Autosomal dominant deafness (DFNA 2), non-insulin dependent diabetes mellitus, CNS disorders, arrhythmia, seizure, asthma, hypertension, development of powerful assay systems for functional analysis of various human disorders which will help in understanding of pathology of the disease, and development of new drug targets for various disorders. They may also be used in diagnostic applications, wherein the presence or amount of the nucleic acid or the protein are to be assessed. These materials are further useful in the generation of antibodies that bind immunospecifically to the novel substances of the invention for use in therapeutic or diagnostic methods.

NOVX Nucleic Acids

The nucleic acids of the invention include those that encode a NOVX polypeptide or protein. As used herein, the terms polypeptide and protein are interchangeable.

5 In some embodiments, a NOVX nucleic acid encodes a mature NOVX polypeptide. As used herein, a "mature" form of a polypeptide or protein described herein relates to the product of a naturally occurring polypeptide or precursor form or proprotein. The naturally occurring polypeptide, precursor or proprotein includes, by way of nonlimiting example, the full-length gene product, encoded by the corresponding gene. Alternatively, it may be defined as the
10 polypeptide, precursor or proprotein encoded by an open reading frame described herein. The product "mature" form arises, again by way of nonlimiting example, as a result of one or more naturally occurring processing steps that may take place within the cell in which the gene product arises. Examples of such processing steps leading to a "mature" form of a polypeptide or protein include the cleavage of the N-terminal methionine residue encoded by the initiation codon of an
15 open reading frame, or the proteolytic cleavage of a signal peptide or leader sequence. Thus a mature form arising from a precursor polypeptide or protein that has residues 1 to N, where residue 1 is the N-terminal methionine, would have residues 2 through N remaining after removal of the N-terminal methionine. Alternatively, a mature form arising from a precursor polypeptide or protein having residues 1 to N, in which an N-terminal signal sequence from
20 residue 1 to residue M is cleaved, would have the residues from residue M+1 to residue N remaining. Further as used herein, a "mature" form of a polypeptide or protein may arise from a step of post-translational modification other than a proteolytic cleavage event. Such additional processes include, by way of non-limiting example, glycosylation, myristoylation or phosphorylation. In general, a mature polypeptide or protein may result from the operation of
25 only one of these processes, or a combination of any of them.

Among the NOVX nucleic acids is the nucleic acid whose sequence is provided in SEQ ID NO: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19 or 21, or a fragment thereof. Additionally, the invention includes mutant or variant nucleic acids of SEQ ID NO: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19 or 21, or a
30 fragment thereof, any of whose bases may be changed from the corresponding bases shown in SEQ ID NO: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19 or 21, while still encoding a protein that maintains at least one of its NOVX-like activities and physiological functions (*i.e.*, modulating angiogenesis, neuronal development). The invention further includes the complement of the nucleic acid

sequence of SEQ ID 1, 3, 5, 7, 9, 11, 13, 15, 17, 19 or 21, including fragments, derivatives, analogs and homologs thereof. The invention additionally includes nucleic acids or nucleic acid fragments, or complements thereto, whose structures include chemical modifications.

One aspect of the invention pertains to isolated nucleic acid molecules that encode NOVX proteins or biologically active portions thereof. Also included are nucleic acid fragments sufficient for use as hybridization probes to identify NOVX-encoding nucleic acids (*e.g.*, NOVX mRNA) and fragments for use as polymerase chain reaction (PCR) primers for the amplification or mutation of NOVX nucleic acid molecules. As used herein, the term "nucleic acid molecule" is intended to include DNA molecules (*e.g.*, cDNA or genomic DNA), RNA molecules (*e.g.*, mRNA), analogs of the DNA or RNA generated using nucleotide analogs, and derivatives, fragments and homologs thereof. The nucleic acid molecule can be single-stranded or double-stranded, but preferably is double-stranded DNA.

"Probes" refer to nucleic acid sequences of variable length, preferably between at least about 10 nucleotides (nt), 100 nt, or as many as about, *e.g.*, 6,000 nt, depending on use. Probes are used in the detection of identical, similar, or complementary nucleic acid sequences. Longer length probes are usually obtained from a natural or recombinant source, are highly specific and much slower to hybridize than oligomers. Probes may be single- or double-stranded and designed to have specificity in PCR, membrane-based hybridization technologies, or ELISA-like technologies.

An "isolated" nucleic acid molecule is one that is separated from other nucleic acid molecules that are present in the natural source of the nucleic acid. Examples of isolated nucleic acid molecules include, but are not limited to, recombinant DNA molecules contained in a vector, recombinant DNA molecules maintained in a heterologous host cell, partially or substantially purified nucleic acid molecules, and synthetic DNA or RNA molecules. Preferably, an "isolated" nucleic acid is free of sequences which naturally flank the nucleic acid (*i.e.*, sequences located at the 5' and 3' ends of the nucleic acid) in the genomic DNA of the organism from which the nucleic acid is derived. For example, in various embodiments, the isolated NOVX nucleic acid molecule can contain less than about 50 kb, 25 kb, 5 kb, 4 kb, 3 kb, 2 kb, 1 kb, 0.5 kb or 0.1 kb of nucleotide sequences which naturally flank the nucleic acid molecule in genomic DNA of the cell from which the nucleic acid is derived. Moreover, an "isolated" nucleic acid molecule, such as a cDNA molecule, can be substantially free of other cellular

material or culture medium when produced by recombinant techniques, or of chemical precursors or other chemicals when chemically synthesized.

A nucleic acid molecule of the present invention, *e.g.*, a nucleic acid molecule having the nucleotide sequence of SEQ ID NO: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19 or 21, or a complement of
5 any of this nucleotide sequence, can be isolated using standard molecular biology techniques and the sequence information provided herein. Using all or a portion of the nucleic acid sequence of SEQ ID NO: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19 or 21, as a hybridization probe, NOVX nucleic acid sequences can be isolated using standard hybridization and cloning techniques (*e.g.*, as described in Sambrook *et al.*, eds., MOLECULAR CLONING: A LABORATORY MANUAL 2nd Ed., Cold Spring
10 Harbor Laboratory Press, Cold Spring Harbor, NY, 1989; and Ausubel, *et al.*, eds., CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, New York, NY, 1993.)

A nucleic acid of the invention can be amplified using cDNA, mRNA or alternatively, genomic DNA, as a template and appropriate oligonucleotide primers according to standard PCR amplification techniques. The nucleic acid so amplified can be cloned into an appropriate vector
15 and characterized by DNA sequence analysis. Furthermore, oligonucleotides corresponding to NOVX nucleotide sequences can be prepared by standard synthetic techniques, *e.g.*, using an automated DNA synthesizer.

As used herein, the term "oligonucleotide" refers to a series of linked nucleotide residues, which oligonucleotide has a sufficient number of nucleotide bases to be used in a PCR reaction.
20 A short oligonucleotide sequence may be based on, or designed from, a genomic or cDNA sequence and is used to amplify, confirm, or reveal the presence of an identical, similar or complementary DNA or RNA in a particular cell or tissue. Oligonucleotides comprise portions of a nucleic acid sequence having about 10 nt, 50 nt, or 100 nt in length, preferably about 15 nt to 30 nt in length. In one embodiment, an oligonucleotide comprising a nucleic acid molecule less
25 than 100 nt in length would further comprise at least 6 contiguous nucleotides of SEQ ID NO: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19 or 21, or a complement thereof. Oligonucleotides may be chemically synthesized and may be used as probes.

In another embodiment, an isolated nucleic acid molecule of the invention comprises a nucleic acid molecule that is a complement of the nucleotide sequence shown in SEQ ID NO: 1,
30 3, 5, 7, 9, 11, 13, 15, 17, 19 or 21, or a portion of this nucleotide sequence. A nucleic acid molecule that is complementary to the nucleotide sequence shown in SEQ ID NO: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19 or 21 is one that is sufficiently complementary to the nucleotide sequence

shown in SEQ ID NO: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19 or 21 that it can hydrogen bond with little or no mismatches to the nucleotide sequence shown in SEQ ID NO: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19 or 21, thereby forming a stable duplex.

As used herein, the term "complementary" refers to Watson-Crick or Hoogsteen base pairing between nucleotide units of a nucleic acid molecule, and the term "binding" means the physical or chemical interaction between two polypeptides or compounds or associated polypeptides or compounds or combinations thereof. Binding includes ionic, non-ionic, Von der Waals, hydrophobic interactions, etc. A physical interaction can be either direct or indirect. Indirect interactions may be through or due to the effects of another polypeptide or compound.

Direct binding refers to interactions that do not take place through, or due to, the effect of another polypeptide or compound, but instead are without other substantial chemical intermediates.

Moreover, the nucleic acid molecule of the invention can comprise only a portion of the nucleic acid sequence of SEQ ID NO: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19 or 21, *e.g.*, a fragment that can be used as a probe or primer, or a fragment encoding a biologically active portion of NOVX.

Fragments provided herein are defined as sequences of at least 6 (contiguous) nucleic acids or at least 4 (contiguous) amino acids, a length sufficient to allow for specific hybridization in the case of nucleic acids or for specific recognition of an epitope in the case of amino acids, respectively, and are at most some portion less than a full length sequence. Fragments may be derived from any contiguous portion of a nucleic acid or amino acid sequence of choice. Derivatives are nucleic acid sequences or amino acid sequences formed from the native compounds either directly or by modification or partial substitution. Analogs are nucleic acid sequences or amino acid sequences that have a structure similar to, but not identical to, the native compound but differs from it in respect to certain components or side chains. Analogs may be synthetic or from a different evolutionary origin and may have a similar or opposite metabolic activity compared to wild type.

Derivatives and analogs may be full length or other than full length, if the derivative or analog contains a modified nucleic acid or amino acid, as described below. Derivatives or analogs of the nucleic acids or proteins of the invention include, but are not limited to, molecules comprising regions that are substantially homologous to the nucleic acids or proteins of the invention, in various embodiments, by at least about 70%, 80%, 85%, 90%, 95%, 98%, or even 99% identity (with a preferred identity of 80-99%) over a nucleic acid or amino acid sequence of identical size or when compared to an aligned sequence in which the alignment is done by a

computer homology program known in the art, or whose encoding nucleic acid is capable of hybridizing to the complement of a sequence encoding the aforementioned proteins under stringent, moderately stringent, or low stringent conditions. See *e.g.* Ausubel, *et al.*, CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, New York, NY, 1993, and below. An exemplary program is the Gap program (Wisconsin Sequence Analysis Package, Version 8 for UNIX, Genetics Computer Group, University Research Park, Madison, WI) using the default settings, which uses the algorithm of Smith and Waterman (Adv. Appl. Math., 1981, 2: 482-489, which is incorporated herein by reference in its entirety).

A "homologous nucleic acid sequence" or "homologous amino acid sequence," or variations thereof, refer to sequences characterized by a homology at the nucleotide level or amino acid level as discussed above. Homologous nucleotide sequences encode those sequences coding for isoforms of a NOVX polypeptide. Isoforms can be expressed in different tissues of the same organism as a result of, for example, alternative splicing of RNA. Alternatively, isoforms can be encoded by different genes. In the present invention, homologous nucleotide sequences include nucleotide sequences encoding for a NOVX polypeptide of species other than humans, including, but not limited to, mammals, and thus can include, *e.g.*, mouse, rat, rabbit, dog, cat, cow, horse, and other organisms. Homologous nucleotide sequences also include, but are not limited to, naturally occurring allelic variations and mutations of the nucleotide sequences set forth herein. A homologous nucleotide sequence does not, however, include the nucleotide sequence encoding human NOVX protein. Homologous nucleic acid sequences include those nucleic acid sequences that encode conservative amino acid substitutions (see below) in SEQ ID NO: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, or 22, as well as a polypeptide having NOVX activity. Biological activities of the NOVX proteins are described below. A homologous amino acid sequence does not encode the amino acid sequence of a human NOVX polypeptide.

The nucleotide sequence determined from the cloning of the human NOVX gene allows for the generation of probes and primers designed for use in identifying and/or cloning NOVX homologues in other cell types, *e.g.*, from other tissues, as well as NOVX homologues from other mammals. The probe/primer typically comprises a substantially purified oligonucleotide. The oligonucleotide typically comprises a region of nucleotide sequence that hybridizes under stringent conditions to at least about 12, 25, 50, 100, 150, 200, 250, 300, 350 or 400 or more consecutive sense strand nucleotide sequence of SEQ ID NO: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19 or

21; or an anti-sense strand nucleotide sequence of SEQ ID NO: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19 or 21; or of a naturally occurring mutant of SEQ ID NO: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19 or 21.

Probes based on the huma NOVX nucleotide sequence can be used to detect transcripts or genomic sequences encoding the same or homologous proteins. In various embodiments, the probe further comprises a label group attached thereto, *e.g.*, the label group can be a radioisotope, a fluorescent compound, an enzyme, or an enzyme co-factor. Such probes can be used as a part of a diagnostic test kit for identifying cells or tissue which misexpress a NOVX protein, such as by measuring a level of a NOVX-encoding nucleic acid in a sample of cells from a subject *e.g.*, detecting NOVX mRNA levels or determining whether a genomic NOVX gene has been mutated or deleted.

A "polypeptide having a biologically active portion of NOVX" refers to polypeptides exhibiting activity similar, but not necessarily identical to, an activity of a polypeptide of the present invention, including mature forms, as measured in a particular biological assay, with or without dose dependency. A nucleic acid fragment encoding a "biologically active portion of NOVX" can be prepared by isolating a portion of SEQ ID NO: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19 or 21 that encodes a polypeptide having a NOVX biological activity (biological activities of the NOVX proteins are described below), expressing the encoded portion of NOVX protein (*e.g.*, by recombinant expression *in vitro*) and assessing the activity of the encoded portion of NOVX. For example, a nucleic acid fragment encoding a biologically active portion of NOVX can optionally include an ATP-binding domain. In another embodiment, a nucleic acid fragment encoding a biologically active portion of NOVX includes one or more regions.

NOVX Variants

The invention further encompasses nucleic acid molecules that differ from the nucleotide sequences shown in SEQ ID NO: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19 or 21 due to the degeneracy of the genetic code. These nucleic acids thus encode the same NOVX protein as that encoded by the nucleotide sequence shown in SEQ ID NO: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19 or 21 *e.g.*, the polypeptide of SEQ ID NO: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, or 22. In another embodiment, an isolated nucleic acid molecule of the invention has a nucleotide sequence encoding a protein having an amino acid sequence shown in SEQ ID NO: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, or 22.

In addition to the huma NOVX nucleotide sequence shown in SEQ ID NO: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19 or 21, it will be appreciated by those skilled in the art that DNA sequence

polymorphisms that lead to changes in the amino acid sequences. NOVX may exist within a population (e.g., the human population). Such genetic polymorphism in the NOVX gene may exist among individuals within a population due to natural allelic variation. As used herein, the terms "gene" and "recombinant gene" refer to nucleic acid molecules comprising an open reading frame encoding a NOVX protein, preferably a mammalia NOVX protein. Such natural allelic variations can typically result in 1-5% variance in the nucleotide sequence of the NOVX gene. Any and all such nucleotide variations and resulting amino acid polymorphisms in NOVX that are the result of natural allelic variation and that do not alter the functional activity of NOVX are intended to be within the scope of the invention.

Moreover, nucleic acid molecules encoding NOVX proteins from other species, and thus that have a nucleotide sequence that differs from the human sequence of SEQ ID NO: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19 or 21 are intended to be within the scope of the invention. Nucleic acid molecules corresponding to natural allelic variants and homologues of the NOVX cDNAs of the invention can be isolated based on their homology to the human NOVX nucleic acids disclosed herein using the human cDNAs, or a portion thereof, as a hybridization probe according to standard hybridization techniques under stringent hybridization conditions. For example, a soluble human NOVX cDNA can be isolated based on its homology to human membrane-bound NOVX. Likewise, a membrane-bound human NOVX cDNA can be isolated based on its homology to soluble human NOVX.

Accordingly, in another embodiment, an isolated nucleic acid molecule of the invention is at least 6 nucleotides in length and hybridizes under stringent conditions to the nucleic acid molecule comprising the nucleotide sequence of SEQ ID NO: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19 or 21. In another embodiment, the nucleic acid is at least 10, 25, 50, 100, 250, 500 or 750 nucleotides in length. In another embodiment, an isolated nucleic acid molecule of the invention hybridizes to the coding region. As used herein, the term "hybridizes under stringent conditions" is intended to describe conditions for hybridization and washing under which nucleotide sequences at least 60% homologous to each other typically remain hybridized to each other.

Homologs (*i.e.*, nucleic acids encoding NOVX proteins derived from species other than human) or other related sequences (*e.g.*, paralogs) can be obtained by low, moderate or high stringency hybridization with all or a portion of the particular human sequence as a probe using methods well known in the art for nucleic acid hybridization and cloning.

As used herein, the phrase "stringent hybridization conditions" refers to conditions under which a probe, primer or oligonucleotide will hybridize to its target sequence, but to no other sequences. Stringent conditions are sequence-dependent and will be different in different circumstances. Longer sequences hybridize specifically at higher temperatures than shorter sequences. Generally, stringent conditions are selected to be about 5°C lower than the thermal melting point (T_m) for the specific sequence at a defined ionic strength and pH. The T_m is the temperature (under defined ionic strength, pH and nucleic acid concentration) at which 50% of the probes complementary to the target sequence hybridize to the target sequence at equilibrium. Since the target sequences are generally present at excess, at T_m , 50% of the probes are occupied at equilibrium. Typically, stringent conditions will be those in which the salt concentration is less than about 1.0 M sodium ion, typically about 0.01 to 1.0 M sodium ion (or other salts) at pH 7.0 to 8.3 and the temperature is at least about 30°C for short probes, primers or oligonucleotides (e.g., 10 nt to 50 nt) and at least about 60°C for longer probes, primers and oligonucleotides. Stringent conditions may also be achieved with the addition of destabilizing agents, such as formamide.

Stringent conditions are known to those skilled in the art and can be found in CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, N.Y. (1989), 6.3.1-6.3.6. Preferably, the conditions are such that sequences at least about 65%, 70%, 75%, 85%, 90%, 95%, 98%, or 99% homologous to each other typically remain hybridized to each other. A non-limiting example of stringent hybridization conditions is hybridization in a high salt buffer comprising 6X SSC, 50 mM Tris-HCl (pH 7.5), 1 mM EDTA, 0.02% PVP, 0.02% Ficoll, 0.02% BSA, and 500 mg/ml denatured salmon sperm DNA at 65°C. This hybridization is followed by one or more washes in 0.2X SSC, 0.01% BSA at 50°C. An isolated nucleic acid molecule of the invention that hybridizes under stringent conditions to the sequence of SEQ ID NO: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19 or 21 corresponds to a naturally occurring nucleic acid molecule. As used herein, a "naturally-occurring" nucleic acid molecule refers to an RNA or DNA molecule having a nucleotide sequence that occurs in nature (e.g., encodes a natural protein).

In a second embodiment, a nucleic acid sequence that is hybridizable to the nucleic acid molecule comprising the nucleotide sequence of SEQ ID NO: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19 or 21, or fragments, analogs or derivatives thereof, under conditions of moderate stringency is provided. A non-limiting example of moderate stringency hybridization conditions are hybridization in 6X SSC, 5X Denhardt's solution, 0.5% SDS and 100 mg/ml denatured salmon

sperm DNA at 55°C, followed by one or more washes in 1X SSC, 0.1% SDS at 37°C. Other conditions of moderate stringency that may be used are well known in the art. See, e.g., Ausubel *et al.* (eds.), 1993, CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, NY, and Kriegler, 1990, GENE TRANSFER AND EXPRESSION, A LABORATORY MANUAL, Stockton Press, NY.

In a third embodiment, a nucleic acid that is hybridizable to the nucleic acid molecule comprising the nucleotide sequence of SEQ ID NO: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19 or 21, or fragments, analogs or derivatives thereof, under conditions of low stringency, is provided. A non-limiting example of low stringency hybridization conditions are hybridization in 35% formamide, 5X SSC, 50 mM Tris-HCl (pH 7.5), 5 mM EDTA, 0.02% PVP, 0.02% Ficoll, 0.2% BSA, 100 mg/ml denatured salmon sperm DNA, 10% (wt/vol) dextran sulfate at 40°C, followed by one or more washes in 2X SSC, 25 mM Tris-HCl (pH 7.4), 5 mM EDTA, and 0.1% SDS at 50°C. Other conditions of low stringency that may be used are well known in the art (e.g., as employed for cross-species hybridizations). See, e.g., Ausubel *et al.* (eds.), 1993, CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, NY, and Kriegler, 1990, GENE TRANSFER AND EXPRESSION, A LABORATORY MANUAL, Stockton Press, NY; Shilo and Weinberg, 1981, *Proc Natl Acad Sci USA* 78: 6789-6792.

Conservative mutations

In addition to naturally-occurring allelic variants of the NOVX sequence that may exist in the population, the skilled artisan will further appreciate that changes can be introduced by mutation into the nucleotide sequence of SEQ ID NO: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19 or 21, thereby leading to changes in the amino acid sequence of the encoded NOVX protein, without altering the functional ability of the NOVX protein. For example, nucleotide substitutions leading to amino acid substitutions at "non-essential" amino acid residues can be made in the sequence of SEQ ID NO: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19 or 21. A "non-essential" amino acid residue is a residue that can be altered from the wild-type sequence of NOVX without altering the biological activity, whereas an "essential" amino acid residue is required for biological activity. For example, amino acid residues that are conserved among the NOVX proteins of the present invention, are predicted to be particularly unamenable to alteration.

Another aspect of the invention pertains to nucleic acid molecules encoding NOVX proteins that contain changes in amino acid residues that are not essential for activity. Such

NOVX proteins differ in amino acid sequence from SEQ ID NO: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, or 22, yet retain biological activity. In one embodiment, the isolated nucleic acid molecule comprises a nucleotide sequence encoding a protein, wherein the protein comprises an amino acid sequence at least about 75% homologous to the amino acid sequence of SEQ ID NO: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, or 22. Preferably, the protein encoded by the nucleic acid is at least about 80% homologous to SEQ ID NO: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, or 22, more preferably at least about 90%, 95%, 98%, and most preferably at least about 99% homologous to SEQ ID NO: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, or 22.

An isolated nucleic acid molecule encoding a NOVX protein homologous to the protein of can be created by introducing one or more nucleotide substitutions, additions or deletions into the nucleotide sequence of SEQ ID NO: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19 or 21, such that one or more amino acid substitutions, additions or deletions are introduced into the encoded protein.

Mutations can be introduced into the nucleotide sequence of SEQ ID NO: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19 or 21 by standard techniques, such as site-directed mutagenesis and PCR-mediated mutagenesis. Preferably, conservative amino acid substitutions are made at one or more predicted non-essential amino acid residues. A "conservative amino acid substitution" is one in which the amino acid residue is replaced with an amino acid residue having a similar side chain. Families of amino acid residues having similar side chains have been defined in the art. These families include amino acids with basic side chains (*e.g.*, lysine, arginine, histidine), acidic side chains (*e.g.*, aspartic acid, glutamic acid), uncharged polar side chains (*e.g.*, glycine, asparagine, glutamine, serine, threonine, tyrosine, cysteine), nonpolar side chains (*e.g.*, alanine, valine, leucine, isoleucine, proline, phenylalanine, methionine, tryptophan), beta-branched side chains (*e.g.*, threonine, valine, isoleucine) and aromatic side chains (*e.g.*, tyrosine, phenylalanine, tryptophan, histidine). Thus, a predicted nonessential amino acid residue in NOVX is replaced with another amino acid residue from the same side chain family. Alternatively, in another embodiment, mutations can be introduced randomly along all or part of a NOVX coding sequence, such as by saturation mutagenesis, and the resultant mutants can be screened for NOVX biological activity to identify mutants that retain activity. Following mutagenesis of SEQ ID NO: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19 or 21 the encoded protein can be expressed by any recombinant technology known in the art and the activity of the protein can be determined.

In one embodiment, a mutant NOVX protein can be assayed for (1) the ability to form protein:protein interactions with other NOVX proteins, other cell-surface proteins, or

biologically active portions thereof; (2) complex formation between a mutant NOVX protein and a NOVX receptor; (3) the ability of a mutant NOVX protein to bind to an intracellular target protein or biologically active portion thereof; (e.g., avidin proteins); (4) the ability to bind NOVX protein; or (5) the ability to specifically bind an anti-NOVX protein antibody.

5

Antisense NOVX Nucleic Acids

Another aspect of the invention pertains to isolated antisense nucleic acid molecules that are hybridizable to or complementary to the nucleic acid molecule comprising the nucleotide sequence of SEQ ID NO: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19 or 21, or fragments, analogs or derivatives thereof. An "antisense" nucleic acid comprises a nucleotide sequence that is complementary to a "sense" nucleic acid encoding a protein, e.g., complementary to the coding strand of a double-stranded cDNA molecule or complementary to an mRNA sequence. In specific aspects, antisense nucleic acid molecules are provided that comprise a sequence complementary to at least about 10, 25, 50, 100, 250 or 500 nucleotides or an entire NOVX coding strand, or to only a portion thereof. Nucleic acid molecules encoding fragments, homologs, derivatives and analogs of a NOVX protein of SEQ ID NO: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, or 22 or antisense nucleic acids complementary to a NOVX nucleic acid sequence of SEQ ID NO: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19 or 21 are additionally provided.

In one embodiment, an antisense nucleic acid molecule is antisense to a "coding region" of the coding strand of a nucleotide sequence encoding NOVX. The term "coding region" refers to the region of the nucleotide sequence comprising codons which are translated into amino acid residues (e.g., the protein coding region of human NOVX corresponds to SEQ ID NO: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, or 22). In another embodiment, the antisense nucleic acid molecule is antisense to a "noncoding region" of the coding strand of a nucleotide sequence encoding NOVX. The term "noncoding region" refers to 5' and 3' sequences which flank the coding region that are not translated into amino acids (i.e., also referred to as 5' and 3' untranslated regions).

Given the coding strand sequences encoding NOVX disclosed herein (e.g., SEQ ID NO: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19 or 21), antisense nucleic acids of the invention can be designed according to the rules of Watson and Crick or Hoogsteen base pairing. The antisense nucleic acid molecule can be complementary to the entire coding region of NOVX mRNA, but more preferably is an oligonucleotide that is antisense to only a portion of the coding or noncoding region of NOVX mRNA. For example, the antisense oligonucleotide can be complementary to

the region surrounding the translation start site of NOVX mRNA. An antisense oligonucleotide can be, for example, about 5, 10, 15, 20, 25, 30, 35, 40, 45 or 50 nucleotides in length. An antisense nucleic acid of the invention can be constructed using chemical synthesis or enzymatic ligation reactions using procedures known in the art. For example, an antisense nucleic acid (e.g., an antisense oligonucleotide) can be chemically synthesized using naturally occurring nucleotides or variously modified nucleotides designed to increase the biological stability of the molecules or to increase the physical stability of the duplex formed between the antisense and sense nucleic acids, e.g., phosphorothioate derivatives and acridine substituted nucleotides can be used.

Examples of modified nucleotides that can be used to generate the antisense nucleic acid include: 5-fluorouracil, 5-bromouracil, 5-chlorouracil, 5-iodouracil, hypoxanthine, xanthine, 4-acetylcytosine, 5-(carboxyhydroxymethyl) uracil, 5-carboxymethylaminomethyl-2-thiouridine, 5-carboxymethylaminomethyluracil, dihydrouracil, beta-D-galactosylqueosine, inosine, N6-isopentenyladenine, 1-methylguanine, 1-methylinosine, 2,2-dimethylguanine, 2-methyladenine, 2-methylguanine, 3-methylcytosine, 5-methylcytosine, N6-adenine, 7-methylguanine, 5-methylaminomethyluracil, 5-methoxyaminomethyl-2-thiouracil, beta-D-mannosylqueosine, 5'-methoxycarboxymethyluracil, 5-methoxyuracil, 2-methylthio-N6-isopentenyladenine, uracil-5-oxyacetic acid (v), wybutoxosine, pseudouracil, queosine, 2-thiocytosine, 5-methyl-2-thiouracil, 2-thiouracil, 4-thiouracil, 5-methyluracil, uracil-5-oxyacetic acid methylester, uracil-5-oxyacetic acid (v), 5-methyl-2-thiouracil, 3-(3-amino-3-N-2-carboxypropyl) uracil, (acp3)w, and 2,6-diaminopurine. Alternatively, the antisense nucleic acid can be produced biologically using an expression vector into which a nucleic acid has been subcloned in an antisense orientation (*i.e.*, RNA transcribed from the inserted nucleic acid will be of an antisense orientation to a target nucleic acid of interest, described further in the following subsection).

The antisense nucleic acid molecules of the invention are typically administered to a subject or generated *in situ* such that they hybridize with or bind to cellular mRNA and/or genomic DNA encoding a NOVX protein to thereby inhibit expression of the protein, e.g., by inhibiting transcription and/or translation. The hybridization can be by conventional nucleotide complementarity to form a stable duplex, or, for example, in the case of an antisense nucleic acid molecule that binds to DNA duplexes, through specific interactions in the major groove of the double helix. An example of a route of administration of antisense nucleic acid molecules of the

invention includes direct injection at a tissue site. Alternatively, antisense nucleic acid molecules can be modified to target selected cells and then administered systemically. For example, for systemic administration, antisense molecules can be modified such that they specifically bind to receptors or antigens expressed on a selected cell surface, *e.g.*, by linking the antisense nucleic acid molecules to peptides or antibodies that bind to cell surface receptors or antigens. The antisense nucleic acid molecules can also be delivered to cells using the vectors described herein. To achieve sufficient intracellular concentrations of antisense molecules, vector constructs in which the antisense nucleic acid molecule is placed under the control of a strong pol II or pol III promoter are preferred.

In yet another embodiment, the antisense nucleic acid molecule of the invention is an α -anomeric nucleic acid molecule. An α -anomeric nucleic acid molecule forms specific double-stranded hybrids with complementary RNA in which, contrary to the usual β -units, the strands run parallel to each other (Gaultier *et al.* (1987) *Nucleic Acids Res* 15: 6625-6641). The antisense nucleic acid molecule can also comprise a 2'-*o*-methylribonucleotide (Inoue *et al.* (1987) *Nucleic Acids Res* 15: 6131-6148) or a chimeric RNA-DNA analogue (Inoue *et al.* (1987) *FEBS Lett* 215: 327-330).

Such modifications include, by way of nonlimiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or derivatized. These modifications are carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject.

NOVX Ribozymes and PNA moieties

In still another embodiment, an antisense nucleic acid of the invention is a ribozyme. Ribozymes are catalytic RNA molecules with ribonuclease activity that are capable of cleaving a single-stranded nucleic acid, such as a mRNA, to which they have a complementary region. Thus, ribozymes (*e.g.*, hammerhead ribozymes (described in Haselhoff and Gerlach (1988) *Nature* 334:585-591)) can be used to catalytically cleave NOVX mRNA transcripts to thereby inhibit translation of NOVX mRNA. A ribozyme having specificity for a NOVX-encoding nucleic acid can be designed based upon the nucleotide sequence of a NOVX DNA disclosed herein (*i.e.*, SEQ ID NO: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19 or 21). For example, a derivative of a *Tetrahymena* L-19 IVS RNA can be constructed in which the nucleotide sequence of the active

site is complementary to the nucleotide sequence to be cleaved in the NOVX-encoding mRNA. See, e.g., Cech *et al.* U.S. Pat. No. 4,987,071; and Cech *et al.* U.S. Pat. No. 5,116,742. Alternatively, NOVX mRNA can be used to select a catalytic RNA having a specific ribonuclease activity from a pool of RNA molecules. See, e.g., Bartel *et al.*, (1993) *Science* 5 261:1411-1418.

Alternatively, NOVX gene expression can be inhibited by targeting nucleotide sequences complementary to the regulatory region of the NOVX (e.g., the NOVX promoter and/or enhancers) to form triple helical structures that prevent transcription of the NOVX gene in target cells. See generally, Helene. (1991) *Anticancer Drug Des.* 6: 569-84; Helene. *et al.* (1992) *Ann.* 10 *N.Y. Acad. Sci.* 660:27-36; and Maher (1992) *Bioassays* 14: 807-15.

In various embodiments, the nucleic acids of NOVX can be modified at the base moiety, sugar moiety or phosphate backbone to improve, e.g., the stability, hybridization, or solubility of the molecule. For example, the deoxyribose phosphate backbone of the nucleic acids can be modified to generate peptide nucleic acids (see Hyrup *et al.* (1996) *Bioorg Med Chem* 4: 5-23). 15 As used herein, the terms "peptide nucleic acids" or "PNAs" refer to nucleic acid mimics, e.g., DNA mimics, in which the deoxyribose phosphate backbone is replaced by a pseudopeptide backbone and only the four natural nucleobases are retained. The neutral backbone of PNAs has been shown to allow for specific hybridization to DNA and RNA under conditions of low ionic strength. The synthesis of PNA oligomers can be performed using standard solid phase peptide 20 synthesis protocols as described in Hyrup *et al.* (1996) above; Perry-O'Keefe *et al.* (1996) *PNAS* 93: 14670-675.

PNAs of NOVX can be used in therapeutic and diagnostic applications. For example, PNAs can be used as antisense or antigene agents for sequence-specific modulation of gene expression by, e.g., inducing transcription or translation arrest or inhibiting replication. PNAs of 25 NOVX can also be used, e.g., in the analysis of single base pair mutations in a gene by, e.g., PNA directed PCR clamping; as artificial restriction enzymes when used in combination with other enzymes, e.g., S1 nucleases (Hyrup B. (1996) above); or as probes or primers for DNA sequence and hybridization (Hyrup *et al.* (1996), above; Perry-O'Keefe (1996), above).

In another embodiment, PNAs of NOVX can be modified, e.g., to enhance their stability 30 or cellular uptake, by attaching lipophilic or other helper groups to PNA, by the formation of PNA-DNA chimeras, or by the use of liposomes or other techniques of drug delivery known in the art. For example, PNA-DNA chimeras of NOVX can be generated that may combine the

advantageous properties of PNA and DNA. Such chimeras allow DNA recognition enzymes, e.g., RNase H and DNA polymerases, to interact with the DNA portion while the PNA portion would provide high binding affinity and specificity. PNA-DNA chimeras can be linked using linkers of appropriate lengths selected in terms of base stacking, number of bonds between the nucleobases, and orientation (Hyrup (1996) above). The synthesis of PNA-DNA chimeras can be performed as described in Hyrup (1996) above and Finn *et al.* (1996) *Nucl Acids Res* 24: 3357-63. For example, a DNA chain can be synthesized on a solid support using standard phosphoramidite coupling chemistry, and modified nucleoside analogs, e.g., 5'-(4-methoxytrityl) amino-5'-deoxy-thymidine phosphoramidite, can be used between the PNA and the 5' end of DNA (Mag *et al.* (1989) *Nucl Acid Res* 17: 5973-88). PNA monomers are then coupled in a stepwise manner to produce a chimeric molecule with a 5' PNA segment and a 3' DNA segment (Finn *et al.* (1996) above). Alternatively, chimeric molecules can be synthesized with a 5' DNA segment and a 3' PNA segment. See, Petersen *et al.* (1975) *Bioorg Med Chem Lett* 5: 1119-1124.

In other embodiments, the oligonucleotide may include other appended groups such as peptides (e.g., for targeting host cell receptors *in vivo*), or agents facilitating transport across the cell membrane (see, e.g., Letsinger *et al.*, 1989, *Proc. Natl. Acad. Sci. U.S.A.* 86:6553-6556; Lemaitre *et al.*, 1987, *Proc. Natl. Acad. Sci.* 84:648-652; PCT Publication No. W088/09810) or the blood-brain barrier (see, e.g., PCT Publication No. W089/10134). In addition, oligonucleotides can be modified with hybridization triggered cleavage agents (See, e.g., Krol *et al.*, 1988, *BioTechniques* 6:958-976) or intercalating agents. (See, e.g., Zon, 1988, *Pharm. Res.* 5: 539-549). To this end, the oligonucleotide may be conjugated to another molecule, e.g., a peptide, a hybridization triggered cross-linking agent, a transport agent, a hybridization-triggered cleavage agent, etc.

25

NOVX Polypeptides

A NOVX polypeptide of the invention includes the NOVX-like protein whose sequence is provided in SEQ ID NO: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, or 22. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residue shown in SEQ ID NO: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, or 22 while still encoding a protein that maintains its NOVX-like activities and physiological functions, or a functional fragment thereof. In some embodiments, up to 20% or more of the residues may be so changed in the mutant or

variant protein. In embodiments, the NOVX polypeptide according to the invention is a mature polypeptide.

In general, a NOVX-like variant that preserves NOVX-like function includes any variant in which residues at a particular position in the sequence have been substituted by other amino acids, and further include the possibility of inserting an additional residue or residues between two residues of the parent protein as well as the possibility of deleting one or more residues from the parent sequence. Any amino acid substitution, insertion, or deletion is encompassed by the invention. In favorable circumstances, the substitution is a conservative substitution as defined above.

One aspect of the invention pertains to isolated NOVX proteins, and biologically active portions thereof, or derivatives, fragments, analogs or homologs thereof. Also provided are polypeptide fragments suitable for use as immunogens to raise anti-NOVX antibodies. In one embodiment, native NOVX proteins can be isolated from cells or tissue sources by an appropriate purification scheme using standard protein purification techniques. In another embodiment, NOVX proteins are produced by recombinant DNA techniques. Alternative to recombinant expression, a NOVX protein or polypeptide can be synthesized chemically using standard peptide synthesis techniques.

An "isolated" or "purified" protein or biologically active portion thereof is substantially free of cellular material or other contaminating proteins from the cell or tissue source from which the NOVX protein is derived, or substantially free from chemical precursors or other chemicals when chemically synthesized. The language "substantially free of cellular material" includes preparations of NOVX protein in which the protein is separated from cellular components of the cells from which it is isolated or recombinantly produced. In one embodiment, the language "substantially free of cellular material" includes preparations of NOVX protein having less than about 30% (by dry weight) of non-NOVX protein (also referred to herein as a "contaminating protein"), more preferably less than about 20% of non-NOVX protein, still more preferably less than about 10% of non-NOVX protein, and most preferably less than about 5% non-NOVX protein. When the NOVX protein or biologically active portion thereof is recombinantly produced, it is also preferably substantially free of culture medium, *i.e.*, culture medium represents less than about 20%, more preferably less than about 10%, and most preferably less than about 5% of the volume of the protein preparation.

The language "substantially free of chemical precursors or other chemicals" includes preparations of NOVX protein in which the protein is separated from chemical precursors or other chemicals that are involved in the synthesis of the protein. In one embodiment, the language "substantially free of chemical precursors or other chemicals" includes preparations of NOVX protein having less than about 30% (by dry weight) of chemical precursors or non-NOVX chemicals, more preferably less than about 20% chemical precursors or non-NOVX chemicals, still more preferably less than about 10% chemical precursors or non-NOVX chemicals, and most preferably less than about 5% chemical precursors or non-NOVX chemicals.

Biologically active portions of a NOVX protein include peptides comprising amino acid sequences sufficiently homologous to or derived from the amino acid sequence of the NOVX protein, *e.g.*, the amino acid sequence shown in SEQ ID NO: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, or 22 that include fewer amino acids than the full length NOVX proteins, and exhibit at least one activity of a NOVX protein. Typically, biologically active portions comprise a domain or motif with at least one activity of the NOVX protein. A biologically active portion of a NOVX protein can be a polypeptide which is, for example, 10, 25, 50, 100 or more amino acids in length.

A biologically active portion of a NOVX protein of the present invention may contain at least one of the above-identified domains conserved between the NOVX proteins, *e.g.* TSR modules. Moreover, other biologically active portions, in which other regions of the protein are deleted, can be prepared by recombinant techniques and evaluated for one or more of the functional activities of a native NOVX protein.

In an embodiment, the NOVX protein has an amino acid sequence shown in SEQ ID NO: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, or 22. In other embodiments, the NOVX protein is substantially homologous to SEQ ID NO: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, or 22 and retains the functional activity of the protein of SEQ ID NO: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, or 22 yet differs in amino acid sequence due to natural allelic variation or mutagenesis, as described in detail below.

Accordingly, in another embodiment, the NOVX protein is a protein that comprises an amino acid sequence at least about 45% homologous to the amino acid sequence of SEQ ID NO: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, or 22 and retains the functional activity of the NOVX proteins of SEQ ID NO: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, or 22.

Determining homology between two or more sequences

To determine the percent homology of two amino acid sequences or of two nucleic acids, the sequences are aligned for optimal comparison purposes (e.g., gaps can be introduced in either of the sequences being compared for optimal alignment between the sequences). The amino acid residues or nucleotides at corresponding amino acid positions or nucleotide positions are then compared. When a position in the first sequence is occupied by the same amino acid residue or nucleotide as the corresponding position in the second sequence, then the molecules are homologous at that position (i.e., as used herein amino acid or nucleic acid "homology" is equivalent to amino acid or nucleic acid "identity").

The nucleic acid sequence homology may be determined as the degree of identity between two sequences. The homology may be determined using computer programs known in the art, such as GAP software provided in the GCG program package. See, *Needleman and Wunsch 1970 J Mol Biol 48: 443-453*. Using GCG GAP software with the following settings for nucleic acid sequence comparison: GAP creation penalty of 5.0 and GAP extension penalty of 0.3, the coding region of the analogous nucleic acid sequences referred to above exhibits a degree of identity preferably of at least 70%, 75%, 80%, 85%, 90%, 95%, 98%, or 99%, with the CDS (encoding) part of the DNA sequence shown in SEQ ID NO: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19 or 21.

The term "sequence identity" refers to the degree to which two polynucleotide or polypeptide sequences are identical on a residue-by-residue basis over a particular region of comparison. The term "percentage of sequence identity" is calculated by comparing two optimally aligned sequences over that region of comparison, determining the number of positions at which the identical nucleic acid base (e.g., A, T, C, G, U, or I, in the case of nucleic acids) occurs in both sequences to yield the number of matched positions, dividing the number of matched positions by the total number of positions in the region of comparison (i.e., the window size), and multiplying the result by 100 to yield the percentage of sequence identity. The term "substantial identity" as used herein denotes a characteristic of a polynucleotide sequence, wherein the polynucleotide comprises a sequence that has at least 80 percent sequence identity, preferably at least 85 percent identity and often 90 to 95 percent sequence identity, more usually at least 99 percent sequence identity as compared to a reference sequence over a comparison region. The term "percentage of positive residues" is calculated by comparing two optimally aligned sequences over that region of comparison, determining the number of positions at which the identical and conservative amino acid substitutions, as defined above, occur in both

sequences to yield the number of matched positions, dividing the number of matched positions by the total number of positions in the region of comparison (*i.e.*, the window size), and multiplying the result by 100 to yield the percentage of positive residues.

5 Chimeric and fusion proteins

The invention also provides NOVX chimeric or fusion proteins. As used herein, a NOVX "chimeric protein" or "fusion protein" comprises a NOVX polypeptide operatively linked to a non-NOVX polypeptide. An "NOVX polypeptide" refers to a polypeptide having an amino acid sequence corresponding to NOVX, whereas a "non-NOVX polypeptide" refers to a polypeptide having an amino acid sequence corresponding to a protein that is not substantially homologous to the NOVX protein, *e.g.*, a protein that is different from the NOVX protein and that is derived from the same or a different organism. Within a NOVX fusion protein the NOVX polypeptide can correspond to all or a portion of a NOVX protein. In one embodiment, a NOVX fusion protein comprises at least one biologically active portion of a NOVX protein. In another embodiment, a NOVX fusion protein comprises at least two biologically active portions of a NOVX protein. Within the fusion protein, the term "operatively linked" is intended to indicate that the NOVX polypeptide and the non-NOVX polypeptide are fused in-frame to each other. The non-NOVX polypeptide can be fused to the N-terminus or C-terminus of the NOVX polypeptide.

For example, in one embodiment a NOVX fusion protein comprises a NOVX polypeptide operably linked to the extracellular domain of a second protein. Such fusion proteins can be further utilized in screening assays for compounds that modulate NOVX activity (such assays are described in detail below).

In another embodiment, the fusion protein is a GST-NOVX fusion protein in which the NOVX sequences are fused to the C-terminus of the GST (*i.e.*, glutathione S-transferase) sequences. Such fusion proteins can facilitate the purification of recombinant NOVX.

In another embodiment, the fusion protein is a NOVX-immunoglobulin fusion protein in which the NOVX sequences comprising one or more domains are fused to sequences derived from a member of the immunoglobulin protein family. The NOVX-immunoglobulin fusion proteins of the invention can be incorporated into pharmaceutical compositions and administered to a subject to inhibit an interaction between a NOVX ligand and a NOVX protein on the surface of a cell, to thereby suppress NOVX-mediated signal transduction *in vivo*. In one nonlimiting

example, a contemplated NOVX ligand of the invention is the NOVX receptor. The NOVX-immunoglobulin fusion proteins can be used to affect the bioavailability of a NOVX cognate ligand. Inhibition of the NOVX ligand/NOVX interaction may be useful therapeutically for both the treatment of proliferative and differentiative disorders, *e.g.*, cancer as well as modulating (*e.g.*, promoting or inhibiting) cell survival, as well as acute and chronic inflammatory disorders and hyperplastic wound healing, *e.g.* hypertrophic scars and keloids. Moreover, the NOVX-immunoglobulin fusion proteins of the invention can be used as immunogens to produce anti-NOVX antibodies in a subject, to purify NOVX ligands, and in screening assays to identify molecules that inhibit the interaction of NOVX with a NOVX ligand.

A NOVX chimeric or fusion protein of the invention can be produced by standard recombinant DNA techniques. For example, DNA fragments coding for the different polypeptide sequences are ligated together in-frame in accordance with conventional techniques, *e.g.*, by employing blunt-ended or stagger-ended termini for ligation, restriction enzyme digestion to provide for appropriate termini, filling-in of cohesive ends as appropriate, alkaline phosphatase treatment to avoid undesirable joining, and enzymatic ligation. In another embodiment, the fusion gene can be synthesized by conventional techniques including automated DNA synthesizers. Alternatively, PCR amplification of gene fragments can be carried out using anchor primers that give rise to complementary overhangs between two consecutive gene fragments that can subsequently be annealed and reamplified to generate a chimeric gene sequence (see, for example, Ausubel et al. (eds.) CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, 1992). Moreover, many expression vectors are commercially available that already encode a fusion moiety (*e.g.*, a GST polypeptide). A NOVX-encoding nucleic acid can be cloned into such an expression vector such that the fusion moiety is linked in-frame to the NOVX protein.

NOVX agonists and antagonists

The present invention also pertains to variants of the NOVX proteins that function as either NOVX agonists (mimetics) or as NOVX antagonists. Variants of the NOVX protein can be generated by mutagenesis, *e.g.*, discrete point mutation or truncation of the NOVX protein.

An agonist of the NOVX protein can retain substantially the same, or a subset of, the biological activities of the naturally occurring form of the NOVX protein. An antagonist of the NOVX protein can inhibit one or more of the activities of the naturally occurring form of the NOVX

protein by, for example, competitively binding to a downstream or upstream member of a cellular signaling cascade which includes the NOVX protein. Thus, specific biological effects can be elicited by treatment with a variant of limited function. In one embodiment, treatment of a subject with a variant having a subset of the biological activities of the naturally occurring form of the protein has fewer side effects in a subject relative to treatment with the naturally occurring form of the NOVX proteins.

Variants of the NOVX protein that function as either NOVX agonists (mimetics) or as NOVX antagonists can be identified by screening combinatorial libraries of mutants, *e.g.*, truncation mutants, of the NOVX protein for NOVX protein agonist or antagonist activity. In one embodiment, a variegated library of NOVX variants is generated by combinatorial mutagenesis at the nucleic acid level and is encoded by a variegated gene library. A variegated library of NOVX variants can be produced by, for example, enzymatically ligating a mixture of synthetic oligonucleotides into gene sequences such that a degenerate set of potential NOVX sequences is expressible as individual polypeptides, or alternatively, as a set of larger fusion proteins (*e.g.*, for phage display) containing the set of NOVX sequences therein. There are a variety of methods which can be used to produce libraries of potential NOVX variants from a degenerate oligonucleotide sequence. Chemical synthesis of a degenerate gene sequence can be performed in an automatic DNA synthesizer, and the synthetic gene then ligated into an appropriate expression vector. Use of a degenerate set of genes allows for the provision, in one mixture, of all of the sequences encoding the desired set of potential NOVX sequences. Methods for synthesizing degenerate oligonucleotides are known in the art (see, *e.g.*, Narang (1983) *Tetrahedron* 39:3; Itakura *et al.* (1984) *Annu Rev Biochem* 53:323; Itakura *et al.* (1984) *Science* 198:1056; Ike *et al.* (1983) *Nucl Acid Res* 11:477).

Polypeptide libraries

In addition, libraries of fragments of the NOVX protein coding sequence can be used to generate a variegated population of NOVX fragments for screening and subsequent selection of variants of a NOVX protein. In one embodiment, a library of coding sequence fragments can be generated by treating a double stranded PCR fragment of a NOVX coding sequence with a nuclease under conditions wherein nicking occurs only about once per molecule, denaturing the double stranded DNA, renaturing the DNA to form double stranded DNA that can include sense/antisense pairs from different nicked products, removing single stranded portions from

reformed duplexes by treatment with S1 nuclease, and ligating the resulting fragment library into an expression vector. By this method, an expression library can be derived which encodes N-terminal and internal fragments of various sizes of the NOVX protein.

Several techniques are known in the art for screening gene products of combinatorial libraries made by point mutations or truncation, and for screening cDNA libraries for gene products having a selected property. Such techniques are adaptable for rapid screening of the gene libraries generated by the combinatorial mutagenesis of NOVX proteins. The most widely used techniques, which are amenable to high throughput analysis, for screening large gene libraries typically include cloning the gene library into replicable expression vectors, transforming appropriate cells with the resulting library of vectors, and expressing the combinatorial genes under conditions in which detection of a desired activity facilitates isolation of the vector encoding the gene whose product was detected. Recursive ensemble mutagenesis (REM), a new technique that enhances the frequency of functional mutants in the libraries, can be used in combination with the screening assays to identify NOVX variants (Arkin and Yourvan (1992) PNAS 89:7811-7815; Delgrave *et al.* (1993) Protein Engineering 6:327-331).

NOVX Antibodies

Also included in the invention are antibodies to NOVX proteins, or fragments of NOVX proteins. The term "antibody" as used herein refers to immunoglobulin molecules and immunologically active portions of immunoglobulin (Ig) molecules, i.e., molecules that contain an antigen binding site that specifically binds (immunoreacts with) an antigen. Such antibodies include, but are not limited to, polyclonal, monoclonal, chimeric, single chain, F_{ab} , $F_{ab'}$, and $F_{(ab)_2}$ fragments, and an F_{ab} expression library. In general, an antibody molecule obtained from humans relates to any of the classes IgG, IgM, IgA, IgE and IgD, which differ from one another by the nature of the heavy chain present in the molecule. Certain classes have subclasses as well, such as IgG₁, IgG₂, and others. Furthermore, in humans, the light chain may be a kappa chain or a lambda chain. Reference herein to antibodies includes a reference to all such classes, subclasses and types of human antibody species.

An isolated NOVX-related protein of the invention may be intended to serve as an antigen, or a portion or fragment thereof, and additionally can be used as an immunogen to generate antibodies that immunospecifically bind the antigen, using standard techniques for polyclonal and monoclonal antibody preparation. The full-length protein can be used or,

alternatively, the invention provides antigenic peptide fragments of the antigen for use as immunogens. An antigenic peptide fragment comprises at least 6 amino acid residues of the amino acid sequence of the full length protein, such as an amino acid sequence shown in from SEQ ID NO: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, or 22, and encompasses an epitope thereof such that an antibody raised against the peptide forms a specific immune complex with the full length protein or with any fragment that contains the epitope. Preferably, the antigenic peptide comprises at least 10 amino acid residues, or at least 15 amino acid residues, or at least 20 amino acid residues, or at least 30 amino acid residues. Preferred epitopes encompassed by the antigenic peptide are regions of the protein that are located on its surface; commonly these are hydrophilic regions.

In certain embodiments of the invention, at least one epitope encompassed by the antigenic peptide is a region of NOVX-related protein that is located on the surface of the protein, *e.g.*, a hydrophilic region. A hydrophobicity analysis of the human NOVX-related protein sequence will indicate which regions of a NOVX-related protein are particularly hydrophilic and, therefore, are likely to encode surface residues useful for targeting antibody production. As a means for targeting antibody production, hydropathy plots showing regions of hydrophilicity and hydrophobicity may be generated by any method well known in the art, including, for example, the Kyte Doolittle or the Hopp Woods methods, either with or without Fourier transformation. See, *e.g.*, Hopp and Woods, 1981, *Proc. Nat. Acad. Sci. USA* 78: 3824-3828; Kyte and Doolittle 1982, *J. Mol. Biol.* 157: 105-142, each of which is incorporated herein by reference in its entirety. Antibodies that are specific for one or more domains within an antigenic protein, or derivatives, fragments, analogs or homologs thereof, are also provided herein.

A protein of the invention, or a derivative, fragment, analog, homolog or ortholog thereof, may be utilized as an immunogen in the generation of antibodies that immunospecifically bind these protein components.

Various procedures known within the art may be used for the production of polyclonal or monoclonal antibodies directed against a protein of the invention, or against derivatives, fragments, analogs homologs or orthologs thereof (see, for example, *Antibodies: A Laboratory Manual*, Harlow E, and Lane D, 1988, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, incorporated herein by reference). Some of these antibodies are discussed below.

P Polyclonal Antibodies

For the production of polyclonal antibodies, various suitable host animals (e.g., rabbit, goat, mouse or other mammal) may be immunized by one or more injections with the native protein, a synthetic variant thereof, or a derivative of the foregoing. An appropriate immunogenic preparation can contain, for example, the naturally occurring immunogenic protein, a chemically synthesized polypeptide representing the immunogenic protein, or a recombinantly expressed immunogenic protein. Furthermore, the protein may be conjugated to a second protein known to be immunogenic in the mammal being immunized. Examples of such immunogenic proteins include but are not limited to keyhole limpet hemocyanin, serum albumin, bovine thyroglobulin, and soybean trypsin inhibitor. The preparation can further include an adjuvant. Various adjuvants used to increase the immunological response include, but are not limited to, Freund's (complete and incomplete), mineral gels (e.g., aluminum hydroxide), surface active substances (e.g., lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, dinitrophenol, etc.), adjuvants usable in humans such as Bacille Calmette-Guerin and Corynebacterium parvum, or similar immunostimulatory agents. Additional examples of adjuvants which can be employed include MPL-TDM adjuvant (monophosphoryl Lipid A, synthetic trehalose dicorynomycolate).

The polyclonal antibody molecules directed against the immunogenic protein can be isolated from the mammal (e.g., from the blood) and further purified by well known techniques, such as affinity chromatography using protein A or protein G, which provide primarily the IgG fraction of immune serum. Subsequently, or alternatively, the specific antigen which is the target of the immunoglobulin sought, or an epitope thereof, may be immobilized on a column to purify the immune specific antibody by immunoaffinity chromatography. Purification of immunoglobulins is discussed, for example, by D. Wilkinson (The Scientist, published by The Scientist, Inc., Philadelphia PA, Vol. 14, No. 8 (April 17, 2000), pp. 25-28).

Monoclonal Antibodies

The term "monoclonal antibody" (MAb) or "monoclonal antibody composition", as used herein, refers to a population of antibody molecules that contain only one molecular species of antibody molecule consisting of a unique light chain gene product and a unique heavy chain gene product. In particular, the complementarity determining regions (CDRs) of the monoclonal antibody are identical in all the molecules of the population. MAbs thus contain an antigen

binding site capable of immunoreacting with a particular epitope of the antigen characterized by a unique binding affinity for it.

Monoclonal antibodies can be prepared using hybridoma methods, such as those described by Kohler and Milstein, Nature, 256:495 (1975). In a hybridoma method, a mouse, hamster, or other appropriate host animal, is typically immunized with an immunizing agent to elicit lymphocytes that produce or are capable of producing antibodies that will specifically bind to the immunizing agent. Alternatively, the lymphocytes can be immunized in vitro.

The immunizing agent will typically include the protein antigen, a fragment thereof or a fusion protein thereof. Generally, either peripheral blood lymphocytes are used if cells of human origin are desired, or spleen cells or lymph node cells are used if non-human mammalian sources are desired. The lymphocytes are then fused with an immortalized cell line using a suitable fusing agent, such as polyethylene glycol, to form a hybridoma cell (Goding, Monoclonal Antibodies: Principles and Practice, Academic Press, (1986) pp. 59-103). Immortalized cell lines are usually transformed mammalian cells, particularly myeloma cells of rodent, bovine and human origin. Usually, rat or mouse myeloma cell lines are employed. The hybridoma cells can be cultured in a suitable culture medium that preferably contains one or more substances that inhibit the growth or survival of the unfused, immortalized cells. For example, if the parental cells lack the enzyme hypoxanthine guanine phosphoribosyl transferase (HGPRT or HPRT), the culture medium for the hybridomas typically will include hypoxanthine, aminopterin, and thymidine ("HAT medium"), which substances prevent the growth of HGPRT-deficient cells.

Preferred immortalized cell lines are those that fuse efficiently, support stable high level expression of antibody by the selected antibody-producing cells, and are sensitive to a medium such as HAT medium. More preferred immortalized cell lines are murine myeloma lines, which can be obtained, for instance, from the Salk Institute Cell Distribution Center, San Diego, California and the American Type Culture Collection, Manassas, Virginia. Human myeloma and mouse-human heteromyeloma cell lines also have been described for the production of human monoclonal antibodies (Kozbor, J. Immunol., 133:3001 (1984); Brodeur et al., Monoclonal Antibody Production Techniques and Applications, Marcel Dekker, Inc., New York, (1987) pp. 51-63).

The culture medium in which the hybridoma cells are cultured can then be assayed for the presence of monoclonal antibodies directed against the antigen. Preferably, the binding specificity of monoclonal antibodies produced by the hybridoma cells is determined by

immunoprecipitation or by an in vitro binding assay, such as radioimmunoassay (RIA) or enzyme-linked immunoabsorbent assay (ELISA). Such techniques and assays are known in the art. The binding affinity of the monoclonal antibody can, for example, be determined by the Scatchard analysis of Munson and Pollard, Anal. Biochem., 107:220 (1980). Preferably, antibodies having a high degree of specificity and a high binding affinity for the target antigen are isolated.

After the desired hybridoma cells are identified, the clones can be subcloned by limiting dilution procedures and grown by standard methods. Suitable culture media for this purpose include, for example, Dulbecco's Modified Eagle's Medium and RPMI-1640 medium.

Alternatively, the hybridoma cells can be grown *in vivo* as ascites in a mammal.

The monoclonal antibodies secreted by the subclones can be isolated or purified from the culture medium or ascites fluid by conventional immunoglobulin purification procedures such as, for example, protein A-Sepharose, hydroxylapatite chromatography, gel electrophoresis, dialysis, or affinity chromatography.

The monoclonal antibodies can also be made by recombinant DNA methods, such as those described in U.S. Patent No. 4,816,567. DNA encoding the monoclonal antibodies of the invention can be readily isolated and sequenced using conventional procedures (e.g., by using oligonucleotide probes that are capable of binding specifically to genes encoding the heavy and light chains of murine antibodies). The hybridoma cells of the invention serve as a preferred source of such DNA. Once isolated, the DNA can be placed into expression vectors, which are then transfected into host cells such as simian COS cells, Chinese hamster ovary (CHO) cells, or myeloma cells that do not otherwise produce immunoglobulin protein, to obtain the synthesis of monoclonal antibodies in the recombinant host cells. The DNA also can be modified, for example, by substituting the coding sequence for human heavy and light chain constant domains in place of the homologous murine sequences (U.S. Patent No. 4,816,567; Morrison, Nature 368, 812-13 (1994)) or by covalently joining to the immunoglobulin coding sequence all or part of the coding sequence for a non-immunoglobulin polypeptide. Such a non-immunoglobulin polypeptide can be substituted for the constant domains of an antibody of the invention, or can be substituted for the variable domains of one antigen-combining site of an antibody of the invention to create a chimeric bivalent antibody.

Humanized Antibodies

transforming human B cells with Epstein Barr Virus in vitro (see C... et al., 1985 In: MONOCLONAL ANTIBODIES AND CANCER THERAPY, Alan R. Liss, Inc., pp. 77-96).

In addition, human antibodies can also be produced using additional techniques, including phage display libraries (Hoogenboom and Winter, J. Mol. Biol., 227:381 (1991); Marks et al., J. Mol. Biol., 222:581 (1991)). Similarly, human antibodies can be made by introducing human immunoglobulin loci into transgenic animals, e.g., mice in which the endogenous immunoglobulin genes have been partially or completely inactivated. Upon challenge, human antibody production is observed, which closely resembles that seen in humans in all respects, including gene rearrangement, assembly, and antibody repertoire. This approach is described, for example, in U.S. Patent Nos. 5,545,807; 5,545,806; 5,569,825; 5,625,126; 5,633,425; 5,661,016, and in Marks et al. (Bio/Technology 10, 779-783 (1992)); Lonberg et al. (Nature 368 856-859 (1994)); Morrison (Nature 368, 812-13 (1994)); Fishwild et al. (Nature Biotechnology 14, 845-51 (1996)); Neuberger (Nature Biotechnology 14, 826 (1996)); and Lonberg and Huszar (Intern. Rev. Immunol. 13 65-93 (1995)).

Human antibodies may additionally be produced using transgenic nonhuman animals which are modified so as to produce fully human antibodies rather than the animal's endogenous antibodies in response to challenge by an antigen. (See PCT publication WO94/02602). The endogenous genes encoding the heavy and light immunoglobulin chains in the nonhuman host have been incapacitated, and active loci encoding human heavy and light chain immunoglobulins are inserted into the host's genome. The human genes are incorporated, for example, using yeast artificial chromosomes containing the requisite human DNA segments. An animal which provides all the desired modifications is then obtained as progeny by crossbreeding intermediate transgenic animals containing fewer than the full complement of the modifications. The preferred embodiment of such a nonhuman animal is a mouse, and is termed the Xenomouse™ as disclosed in PCT publications WO 96/33735 and WO 96/34096. This animal produces B cells which secrete fully human immunoglobulins. The antibodies can be obtained directly from the animal after immunization with an immunogen of interest, as, for example, a preparation of a polyclonal antibody, or alternatively from immortalized B cells derived from the animal, such as hybridomas producing monoclonal antibodies. Additionally, the genes encoding the immunoglobulins with human variable regions can be recovered and expressed to obtain the antibodies directly, or can be further modified to obtain analogs of antibodies such as, for example, single chain Fv molecules.

An example method of producing a nonhuman host, exemplified as a mouse, lacking expression of an endogenous immunoglobulin heavy chain is disclosed in U.S. Patent No. 5,939,598. It can be obtained by a method including deleting the J segment genes from at least one endogenous heavy chain locus in an embryonic stem cell to prevent rearrangement of the locus and to prevent formation of a transcript of a rearranged immunoglobulin heavy chain locus, the deletion being effected by a targeting vector containing a gene encoding a selectable marker; and producing from the embryonic stem cell a transgenic mouse whose somatic and germ cells contain the gene encoding the selectable marker.

A method for producing an antibody of interest, such as a human antibody, is disclosed in U.S. Patent No. 5,916,771. It includes introducing an expression vector that contains a nucleotide sequence encoding a heavy chain into one mammalian host cell in culture, introducing an expression vector containing a nucleotide sequence encoding a light chain into another mammalian host cell, and fusing the two cells to form a hybrid cell. The hybrid cell expresses an antibody containing the heavy chain and the light chain.

In a further improvement on this procedure, a method for identifying a clinically relevant epitope on an immunogen, and a correlative method for selecting an antibody that binds immunospecifically to the relevant epitope with high affinity, are disclosed in PCT publication WO 99/53049.

F_{ab} Fragments and Single Chain Antibodies

According to the invention, techniques can be adapted for the production of single-chain antibodies specific to an antigenic protein of the invention (see e.g., U.S. Patent No. 4,946,778). In addition, methods can be adapted for the construction of F_{ab} expression libraries (see e.g., Huse, et al., 1989 Science 246: 1275-1281) to allow rapid and effective identification of monoclonal F_{ab} fragments with the desired specificity for a protein or derivatives, fragments, analogs or homologs thereof. Antibody fragments that contain the idiotype to a protein antigen may be produced by techniques known in the art including, but not limited to: (i) an F_{(ab)2} fragment produced by pepsin digestion of an antibody molecule; (ii) an F_{ab} fragment generated by reducing the disulfide bridges of an F_{(ab)2} fragment; (iii) an F_{ab} fragment generated by the treatment of the antibody molecule with papain and a reducing agent and (iv) F_v fragments.

Bispecific Antibodies

Bispecific antibodies are monoclonal, preferably human or humanized, antibodies that have binding specificities for at least two different antigens. In the present case, one of the binding specificities is for an antigenic protein of the invention. The second binding target is any other antigen, and advantageously is a cell-surface protein or receptor or receptor subunit.

5 Methods for making bispecific antibodies are known in the art. Traditionally, the recombinant production of bispecific antibodies is based on the co-expression of two immunoglobulin heavy-chain/light-chain pairs, where the two heavy chains have different specificities (Milstein and Cuello, Nature, 305:537-539 (1983)). Because of the random assortment of immunoglobulin heavy and light chains, these hybridomas (quadromas) produce a
10 potential mixture of ten different antibody molecules, of which only one has the correct bispecific structure. The purification of the correct molecule is usually accomplished by affinity chromatography steps. Similar procedures are disclosed in WO 93/08829, published 13 May 1993, and in Traunecker *et al.*, 1991 *EMBO J.*, 10:3655-3659.

Antibody variable domains with the desired binding specificities (antibody-antigen
15 combining sites) can be fused to immunoglobulin constant domain sequences. The fusion preferably is with an immunoglobulin heavy-chain constant domain, comprising at least part of the hinge, CH2, and CH3 regions. It is preferred to have the first heavy-chain constant region (CH1) containing the site necessary for light-chain binding present in at least one of the fusions. DNAs encoding the immunoglobulin heavy-chain fusions and, if desired, the immunoglobulin
20 light chain, are inserted into separate expression vectors, and are co-transfected into a suitable host organism. For further details of generating bispecific antibodies see, for example, Suresh *et al.*, Methods in Enzymology, 121:210 (1986).

According to another approach described in WO 96/27011, the interface between a pair
25 of antibody molecules can be engineered to maximize the percentage of heterodimers which are recovered from recombinant cell culture. The preferred interface comprises at least a part of the CH3 region of an antibody constant domain. In this method, one or more small amino acid side chains from the interface of the first antibody molecule are replaced with larger side chains (e.g. tyrosine or tryptophan). Compensatory "cavities" of identical or similar size to the large side chain(s) are created on the interface of the second antibody molecule by replacing large amino
30 acid side chains with smaller ones (e.g. alanine or threonine). This provides a mechanism for increasing the yield of the heterodimer over other unwanted end-products such as homodimers.

Bispecific antibodies can be prepared as full length antibodies or antibody fragments (e.g. $F(ab')_2$ bispecific antibodies). Techniques for generating bispecific antibodies from antibody fragments have been described in the literature. For example, bispecific antibodies can be prepared using chemical linkage. Brennan et al., Science 229:81 (1985) describe a procedure wherein intact antibodies are proteolytically cleaved to generate $F(ab')_2$ fragments. These fragments are reduced in the presence of the dithiol complexing agent sodium arsenite to stabilize vicinal dithiols and prevent intermolecular disulfide formation. The Fab' fragments generated are then converted to thionitrobenzoate (TNB) derivatives. One of the Fab'-TNB derivatives is then reconverted to the Fab'-thiol by reduction with mercaptoethylamine and is mixed with an equimolar amount of the other Fab'-TNB derivative to form the bispecific antibody. The bispecific antibodies produced can be used as agents for the selective immobilization of enzymes.

Additionally, Fab' fragments can be directly recovered from E. coli and chemically coupled to form bispecific antibodies. Shalaby et al., J. Exp. Med. 175:217-225 (1992) describe the production of a fully humanized bispecific antibody $F(ab')_2$ molecule. Each Fab' fragment was separately secreted from E. coli and subjected to directed chemical coupling in vitro to form the bispecific antibody. The bispecific antibody thus formed was able to bind to cells overexpressing the ErbB2 receptor and normal human T cells, as well as trigger the lytic activity of human cytotoxic lymphocytes against human breast tumor targets.

Various techniques for making and isolating bispecific antibody fragments directly from recombinant cell culture have also been described. For example, bispecific antibodies have been produced using leucine zippers. Kostelny et al., J. Immunol. 148(5):1547-1553 (1992). The leucine zipper peptides from the Fos and Jun proteins were linked to the Fab' portions of two different antibodies by gene fusion. The antibody homodimers were reduced at the hinge region to form monomers and then re-oxidized to form the antibody heterodimers. This method can also be utilized for the production of antibody homodimers. The "diabody" technology described by Hollinger et al., Proc. Natl. Acad. Sci. USA 90:6444-6448 (1993) has provided an alternative mechanism for making bispecific antibody fragments. The fragments comprise a heavy-chain variable domain (V_H) connected to a light-chain variable domain (V_L) by a linker which is too short to allow pairing between the two domains on the same chain. Accordingly, the V_H and V_L domains of one fragment are forced to pair with the complementary V_L and V_H domains of another fragment, thereby forming two antigen-binding sites. Another strategy for making

bispecific antibody fragments by the use of single-chain Fv (sFv) has also been reported. See, Gruber et al., J. Immunol. 152:5368 (1994).

Antibodies with more than two valencies are contemplated. For example, trispecific antibodies can be prepared. Tutt et al., J. Immunol. 147:60 (1991).

5 Exemplary bispecific antibodies can bind to two different epitopes, at least one of which originates in the protein antigen of the invention. Alternatively, an anti-antigenic arm of an immunoglobulin molecule can be combined with an arm which binds to a triggering molecule on a leukocyte such as a T-cell receptor molecule (e.g. CD2, CD3, CD28, or B7), or Fc receptors for IgG (Fc R), such as Fc-RI (CD64), Fc RII (CD32) and Fc RIII (CD16) so as to focus cellular
10 defense mechanisms to the cell expressing the particular antigen. Bispecific antibodies can also be used to direct cytotoxic agents to cells which express a particular antigen. These antibodies possess an antigen-binding arm and an arm which binds a cytotoxic agent or a radionuclide chelator, such as EOTUBE, DPTA, DOTA, or TETA. Another bispecific antibody of interest binds the protein antigen described herein and further binds tissue factor (TF).

15

Heteroconjugate Antibodies

Heteroconjugate antibodies are also within the scope of the present invention. Heteroconjugate antibodies are composed of two covalently joined antibodies. Such antibodies have, for example, been proposed to target immune system cells to unwanted cells (U.S. Patent
20 No. 4,676,980), and for treatment of HIV infection (WO 91/00360; WO 92/200373; EP 03089). It is contemplated that the antibodies can be prepared in vitro using known methods in synthetic protein chemistry, including those involving crosslinking agents. For example, immunotoxins can be constructed using a disulfide exchange reaction or by forming a thioether bond. Examples of suitable reagents for this purpose include iminothiolate and methyl-4-mercaptobutyrimidate
25 and those disclosed, for example, in U.S. Patent No. 4,676,980.

Effector Function Engineering

It can be desirable to modify the antibody of the invention with respect to effector function, so as to enhance, e.g., the effectiveness of the antibody in treating cancer. For example,
30 cysteine residue(s) can be introduced into the Fc region, thereby allowing interchain disulfide bond formation in this region. The homodimeric antibody thus generated can have improved internalization capability and/or increased complement-mediated cell killing and antibody-

dependent cellular cytotoxicity (ADCC). See Caron et al., J. Exp. Med., 176: 1191-1195 (1992) and Shopes, J. Immunol., 148: 2918-2922 (1992). Homodimeric antibodies with enhanced anti-tumor activity can also be prepared using heterobifunctional cross-linkers as described in Wolff et al. Cancer Research, 53: 2560-2565 (1993). Alternatively, an antibody can be engineered that has dual Fc regions and can thereby have enhanced complement lysis and ADCC capabilities. See Stevenson et al., Anti-Cancer Drug Design, 3: 219-230 (1989).

Immunoconjugates

The invention also pertains to immunoconjugates comprising an antibody conjugated to a cytotoxic agent such as a chemotherapeutic agent, toxin (e.g., an enzymatically active toxin of bacterial, fungal, plant, or animal origin, or fragments thereof), or a radioactive isotope (i.e., a radioconjugate).

Chemotherapeutic agents useful in the generation of such immunoconjugates have been described above. Enzymatically active toxins and fragments thereof that can be used include diphtheria A chain, nonbinding active fragments of diphtheria toxin, exotoxin A chain (from *Pseudomonas aeruginosa*), ricin A chain, abrin A chain, modeccin A chain, alpha-sarcin, Aleurites fordii proteins, dianthin proteins, Phytolacca americana proteins (PAPI, PAPII, and PAP-S), momordica charantia inhibitor, curcin, crotin, sapaonaria officinalis inhibitor, gelonin, mitogellin, restrictocin, phenomycin, enomycin, and the tricothecenes. A variety of radionuclides are available for the production of radioconjugated antibodies. Examples include ^{212}Bi , ^{131}I , ^{131}In , ^{90}Y , and ^{186}Re .

Conjugates of the antibody and cytotoxic agent are made using a variety of bifunctional protein-coupling agents such as N-succinimidyl-3-(2-pyridyldithiol) propionate (SPDP), iminothiolane (IT), bifunctional derivatives of imidoesters (such as dimethyl adipimidate HCL), active esters (such as disuccinimidyl suberate), aldehydes (such as glutaraldehyde), bis-azido compounds (such as bis (p-azidobenzoyl) hexanediamine), bis-diazonium derivatives (such as bis-(p-diazoniumbenzoyl)-ethylenediamine), diisocyanates (such as tolyene 2,6-diisocyanate), and bis-active fluorine compounds (such as 1,5-difluoro-2,4-dinitrobenzene). For example, a ricin immunotoxin can be prepared as described in Vitetta et al., Science, 238: 1098 (1987). Carbon-14-labeled 1-isothiocyanatobenzyl-3-methyldiethylene triaminepentaacetic acid (MX-DTPA) is an exemplary chelating agent for conjugation of radionucleotide to the antibody. See WO94/11026.

In another embodiment, the antibody can be conjugated to a "receptor" (such streptavidin) for utilization in tumor pretargeting wherein the antibody-receptor conjugate is administered to the patient, followed by removal of unbound conjugate from the circulation using a clearing agent and then administration of a "ligand" (e.g., avidin) that is in turn conjugated to a cytotoxic agent.

NOVX Recombinant Expression Vectors and Host Cells

Another aspect of the invention pertains to vectors, preferably expression vectors, containing a nucleic acid encoding a NOVX protein, or derivatives, fragments, analogs or homologs thereof. As used herein, the term "vector" refers to a nucleic acid molecule capable of transporting another nucleic acid to which it has been linked. One type of vector is a "plasmid", which refers to a circular double stranded DNA loop into which additional DNA segments can be ligated. Another type of vector is a viral vector, wherein additional DNA segments can be ligated into the viral genome. Certain vectors are capable of autonomous replication in a host cell into which they are introduced (e.g., bacterial vectors having a bacterial origin of replication and episomal mammalian vectors). Other vectors (e.g., non-episomal mammalian vectors) are integrated into the genome of a host cell upon introduction into the host cell, and thereby are replicated along with the host genome. Moreover, certain vectors are capable of directing the expression of genes to which they are operatively-linked. Such vectors are referred to herein as "expression vectors". In general, expression vectors of utility in recombinant DNA techniques are often in the form of plasmids. In the present specification, "plasmid" and "vector" can be used interchangeably as the plasmid is the most commonly used form of vector. However, the invention is intended to include such other forms of expression vectors, such as viral vectors (e.g., replication defective retroviruses, adenoviruses and adeno-associated viruses), which serve equivalent functions.

The recombinant expression vectors of the invention comprise a nucleic acid of the invention in a form suitable for expression of the nucleic acid in a host cell, which means that the recombinant expression vectors include one or more regulatory sequences, selected on the basis of the host cells to be used for expression, that is operatively-linked to the nucleic acid sequence to be expressed. Within a recombinant expression vector, "operably-linked" is intended to mean that the nucleotide sequence of interest is linked to the regulatory sequence(s) in a manner that

allows for expression of the nucleotide sequence (e.g., in an *in vitro* transcription/translation system or in a host cell when the vector is introduced into the host cell).

The term "regulatory sequence" is intended to include promoters, enhancers and other expression control elements (e.g., polyadenylation signals). Such regulatory sequences are described, for example, in Goeddel, GENE EXPRESSION TECHNOLOGY: METHODS IN ENZYMOLOGY 185, Academic Press, San Diego, Calif. (1990). Regulatory sequences include those that direct constitutive expression of a nucleotide sequence in many types of host cell and those that direct expression of the nucleotide sequence only in certain host cells (e.g., tissue-specific regulatory sequences). It will be appreciated by those skilled in the art that the design of the expression vector can depend on such factors as the choice of the host cell to be transformed, the level of expression of protein desired, etc. The expression vectors of the invention can be introduced into host cells to thereby produce proteins or peptides, including fusion proteins or peptides, encoded by nucleic acids as described herein (e.g., NOVX proteins, mutant forms of NOVX proteins, fusion proteins, etc.).

The recombinant expression vectors of the invention can be designed for expression of NOVX proteins in prokaryotic or eukaryotic cells. For example, NOVX proteins can be expressed in bacterial cells such as *Escherichia coli*, insect cells (using baculovirus expression vectors) yeast cells or mammalian cells. Suitable host cells are discussed further in Goeddel, GENE EXPRESSION TECHNOLOGY: METHODS IN ENZYMOLOGY 185, Academic Press, San Diego, Calif. (1990). Alternatively, the recombinant expression vector can be transcribed and translated *in vitro*, for example using T7 promoter regulatory sequences and T7 polymerase.

Expression of proteins in prokaryotes is most often carried out in *Escherichia coli* with vectors containing constitutive or inducible promoters directing the expression of either fusion or non-fusion proteins. Fusion vectors add a number of amino acids to a protein encoded therein, usually to the amino terminus of the recombinant protein. Such fusion vectors typically serve three purposes: (i) to increase expression of recombinant protein; (ii) to increase the solubility of the recombinant protein; and (iii) to aid in the purification of the recombinant protein by acting as a ligand in affinity purification. Often, in fusion expression vectors, a proteolytic cleavage site is introduced at the junction of the fusion moiety and the recombinant protein to enable separation of the recombinant protein from the fusion moiety subsequent to purification of the fusion protein. Such enzymes, and their cognate recognition sequences, include Factor Xa, thrombin and enterokinase. Typical fusion expression vectors include pGEX (Pharmacia Biotech

Inc; Smith and Johnson, 1988. *Gene* 67: 31-40), pMAL (New England Biolabs, Beverly, Mass.) and pRIT5 (Pharmacia, Piscataway, N.J.) that fuse glutathione S-transferase (GST), maltose E binding protein, or protein A, respectively, to the target recombinant protein.

5 Examples of suitable inducible non-fusion *E. coli* expression vectors include pTrc (Amrann *et al.*, (1988) *Gene* 69:301-315) and pET 11d (Studier *et al.*, GENE EXPRESSION TECHNOLOGY: METHODS IN ENZYMOLOGY 185, Academic Press, San Diego, Calif. (1990) 60-89).

10 One strategy to maximize recombinant protein expression in *E. coli* is to express the protein in a host bacteria with an impaired capacity to proteolytically cleave the recombinant protein. See, e.g., Gottesman, GENE EXPRESSION TECHNOLOGY: METHODS IN ENZYMOLOGY 185, Academic Press, San Diego, Calif. (1990) 119-128. Another strategy is to alter the nucleic acid sequence of the nucleic acid to be inserted into an expression vector so that the individual codons for each amino acid are those preferentially utilized in *E. coli* (see, e.g., Wada, *et al.*, 1992. *Nucl. Acids Res.* 20: 2111-2118). Such alteration of nucleic acid sequences of the invention can be
15 carried out by standard DNA synthesis techniques.

In another embodiment, the NOVX expression vector is a yeast expression vector. Examples of vectors for expression in yeast *Saccharomyces cerevisiae* include pYepSec1 (Baldari, *et al.*, 1987. *EMBO J.* 6: 229-234), pMFa (Kurjan and Herskowitz, 1982. *Cell* 30: 933-943), pJRY88 (Schultz *et al.*, 1987. *Gene* 54: 113-123), pYES2 (Invitrogen Corporation,
20 San Diego, Calif.), and picZ (InVitrogen Corp, San Diego, Calif.).

Alternatively, NOVX can be expressed in insect cells using baculovirus expression vectors. Baculovirus vectors available for expression of proteins in cultured insect cells (e.g., SF9 cells) include the pAc series (Smith, *et al.*, 1983. *Mol. Cell. Biol.* 3: 2156-2165) and the pVL series (Lucklow and Summers, 1989. *Virology* 170: 31-39).

25 In yet another embodiment, a nucleic acid of the invention is expressed in mammalian cells using a mammalian expression vector. Examples of mammalian expression vectors include pCDM8 (Seed, 1987. *Nature* 329: 840) and pMT2PC (Kaufman, *et al.*, 1987. *EMBO J.* 6: 187-195). When used in mammalian cells, the expression vector's control functions are often provided by viral regulatory elements. For example, commonly used promoters are derived from
30 polyoma, adenovirus 2, cytomegalovirus, and simian virus 40. For other suitable expression systems for both prokaryotic and eukaryotic cells see, e.g., Chapters 16 and 17 of Sambrook, *et*

al., MOLECULAR CLONING: A LABORATORY MANUAL. 2nd ed., Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1989.

In another embodiment, the recombinant mammalian expression vector is capable of directing expression of the nucleic acid preferentially in a particular cell type (e.g., tissue-specific regulatory elements are used to express the nucleic acid). Tissue-specific regulatory elements are known in the art. Non-limiting examples of suitable tissue-specific promoters include the albumin promoter (liver-specific; Pinkert, *et al.*, 1987. *Genes Dev.* 1: 268-277), lymphoid-specific promoters (Calame and Eaton, 1988. *Adv. Immunol.* 43: 235-275), in particular promoters of T cell receptors (Winoto and Baltimore, 1989. *EMBO J.* 8: 729-733) and immunoglobulins (Banerji, *et al.*, 1983. *Cell* 33: 729-740; Queen and Baltimore, 1983. *Cell* 33: 741-748), neuron-specific promoters (e.g., the neurofilament promoter; Byrne and Ruddle, 1989. *Proc. Natl. Acad. Sci. USA* 86: 5473-5477), pancreas-specific promoters (Edlund, *et al.*, 1985. *Science* 230: 912-916), and mammary gland-specific promoters (e.g., milk whey promoter; U.S. Pat. No. 4,873,316 and European Application Publication No. 264,166). Developmentally-regulated promoters are also encompassed, e.g., the murine hox promoters (Kessel and Gruss, 1990. *Science* 249: 374-379) and the α -fetoprotein promoter (Campes and Tilghman, 1989. *Genes Dev.* 3: 537-546).

The invention further provides a recombinant expression vector comprising a DNA molecule of the invention cloned into the expression vector in an antisense orientation. That is, the DNA molecule is operatively-linked to a regulatory sequence in a manner that allows for expression (by transcription of the DNA molecule) of an RNA molecule that is antisense to NOVX mRNA. Regulatory sequences operatively linked to a nucleic acid cloned in the antisense orientation can be chosen that direct the continuous expression of the antisense RNA molecule in a variety of cell types, for instance viral promoters and/or enhancers, or regulatory sequences can be chosen that direct constitutive, tissue specific or cell type specific expression of antisense RNA. The antisense expression vector can be in the form of a recombinant plasmid, phagemid or attenuated virus in which antisense nucleic acids are produced under the control of a high efficiency regulatory region, the activity of which can be determined by the cell type into which the vector is introduced. For a discussion of the regulation of gene expression using antisense genes see, e.g., Weintraub, *et al.*, "Antisense RNA as a molecular tool for genetic analysis," *Reviews-Trends in Genetics*, Vol. 1(1) 1986.

Another aspect of the invention pertains to host cells into which a recombinant

expression vector of the invention has been introduced. The terms "host cell" and "recombinant host cell" are used interchangeably herein. It is understood that such terms refer not only to the particular subject cell but also to the progeny or potential progeny of such a cell. Because certain modifications may occur in succeeding generations due to either mutation or environmental influences, such progeny may not, in fact, be identical to the parent cell, but are still included within the scope of the term as used herein.

A host cell can be any prokaryotic or eukaryotic cell. For example, NOVX protein can be expressed in bacterial cells such as *E. coli*, insect cells, yeast or mammalian cells (such as human, Chinese hamster ovary cells (CHO) or COS cells). Other suitable host cells are known to those skilled in the art.

Vector DNA can be introduced into prokaryotic or eukaryotic cells via conventional transformation or transfection techniques. As used herein, the terms "transformation" and "transfection" are intended to refer to a variety of art-recognized techniques for introducing foreign nucleic acid (*e.g.*, DNA) into a host cell, including calcium phosphate or calcium chloride co-precipitation, DEAE-dextran-mediated transfection, lipofection, or electroporation. Suitable methods for transforming or transfecting host cells can be found in Sambrook, *et al.* (MOLECULAR CLONING: A LABORATORY MANUAL. 2nd ed., Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1989), and other laboratory manuals.

For stable transfection of mammalian cells, it is known that, depending upon the expression vector and transfection technique used, only a small fraction of cells may integrate the foreign DNA into their genome. In order to identify and select these integrants, a gene that encodes a selectable marker (*e.g.*, resistance to antibiotics) is generally introduced into the host cells along with the gene of interest. Various selectable markers include those that confer resistance to drugs, such as G418, hygromycin and methotrexate. Nucleic acid encoding a selectable marker can be introduced into a host cell on the same vector as that encoding NOVX or can be introduced on a separate vector. Cells stably transfected with the introduced nucleic acid can be identified by drug selection (*e.g.*, cells that have incorporated the selectable marker gene will survive, while the other cells die).

A host cell of the invention, such as a prokaryotic or eukaryotic host cell in culture, can be used to produce (*i.e.*, express) NOVX protein. Accordingly, the invention further provides methods for producing NOVX protein using the host cells of the invention. In one embodiment,

the method comprises culturing the host cell of invention (into which a recombinant expression vector encoding NOVX protein has been introduced) in a suitable medium such that NOVX protein is produced. In another embodiment, the method further comprises isolating NOVX protein from the medium or the host cell.

5

Transgenic NOVX Animals

The host cells of the invention can also be used to produce non-human transgenic animals. For example, in one embodiment, a host cell of the invention is a fertilized oocyte or an embryonic stem cell into which NOVX protein-coding sequences have been introduced. Such
10 host cells can then be used to create non-human transgenic animals in which exogenous NOVX sequences have been introduced into their genome or homologous recombinant animals in which endogenous NOVX sequences have been altered. Such animals are useful for studying the function and/or activity of NOVX protein and for identifying and/or evaluating modulators of NOVX protein activity. As used herein, a "transgenic animal" is a non-human animal, preferably
15 a mammal, more preferably a rodent such as a rat or mouse, in which one or more of the cells of the animal includes a transgene. Other examples of transgenic animals include non-human primates, sheep, dogs, cows, goats, chickens, amphibians, etc. A transgene is exogenous DNA that is integrated into the genome of a cell from which a transgenic animal develops and that remains in the genome of the mature animal, thereby directing the expression of an encoded gene
20 product in one or more cell types or tissues of the transgenic animal. As used herein, a "homologous recombinant animal" is a non-human animal, preferably a mammal, more preferably a mouse, in which an endogenous NOVX gene has been altered by homologous recombination between the endogenous gene and an exogenous DNA molecule introduced into a cell of the animal, *e.g.*, an embryonic cell of the animal, prior to development of the animal.

25 A transgenic animal of the invention can be created by introducing NOVX-encoding nucleic acid into the male pronuclei of a fertilized oocyte (*e.g.*, by microinjection, retroviral infection) and allowing the oocyte to develop in a pseudopregnant female foster animal. Sequences including SEQ ID NO: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19 or 21 can be introduced as a transgene into the genome of a non-human animal. Alternatively, a non-human homologue of
30 the huma NOVX gene, such as a mouse NOVX gene, can be isolated based on hybridization to the huma NOVX cDNA (described further *supra*) and used as a transgene. Intronic sequences and polyadenylation signals can also be included in the transgene to increase the efficiency of

expression of the transgene. A tissue-specific regulatory sequence(s) can be operably-linked to the NOVX transgene to direct expression of NOVX protein to particular cells. Methods for generating transgenic animals via embryo manipulation and microinjection, particularly animals such as mice, have become conventional in the art and are described, for example, in U.S. Patent Nos. 4,736,866; 4,870,009; and 4,873,191; and Hogan, 1986. In: MANIPULATING THE MOUSE EMBRYO, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. Similar methods are used for production of other transgenic animals. A transgenic founder animal can be identified based upon the presence of the NOVX transgene in its genome and/or expression of NOVX mRNA in tissues or cells of the animals. A transgenic founder animal can then be used to breed additional animals carrying the transgene. Moreover, transgenic animals carrying a transgene-encoding NOVX protein can further be bred to other transgenic animals carrying other transgenes.

To create a homologous recombinant animal, a vector is prepared which contains at least a portion of a NOVX gene into which a deletion, addition or substitution has been introduced to thereby alter, *e.g.*, functionally disrupt, the NOVX gene. The NOVX gene can be a human gene (*e.g.*, the DNA of SEQ ID NO: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19 or 21), but more preferably, is a non-human homologue of a human NOVX gene. For example, a mouse homologue of human NOVX gene of SEQ ID NO: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19 or 21 can be used to construct a homologous recombination vector suitable for altering an endogenous NOVX gene in the mouse genome. In one embodiment, the vector is designed such that, upon homologous recombination, the endogenous NOVX gene is functionally disrupted (*i.e.*, no longer encodes a functional protein; also referred to as a "knock out" vector).

Alternatively, the vector can be designed such that, upon homologous recombination, the endogenous NOVX gene is mutated or otherwise altered but still encodes functional protein (*e.g.*, the upstream regulatory region can be altered to thereby alter the expression of the endogenous NOVX protein). In the homologous recombination vector, the altered portion of the NOVX gene is flanked at its 5'- and 3'-termini by additional nucleic acid of the NOVX gene to allow for homologous recombination to occur between the exogenous NOVX gene carried by the vector and an endogenous NOVX gene in an embryonic stem cell. The additional flanking NOVX nucleic acid is of sufficient length for successful homologous recombination with the endogenous gene. Typically, several kilobases of flanking DNA (both at the 5'- and 3'-termini) are included in the vector. *See, e.g.*, Thomas, *et al.*, 1987. *Cell* 51: 503 for a description of

homologous recombination vectors. The vector is then introduced into an embryonic stem cell line (e.g., by electroporation) and cells in which the introduced NOVX gene has homologously-recombined with the endogenous NOVX gene are selected. See, e.g., Li, *et al.*, 1992. *Cell* 69: 915.

5 The selected cells are then injected into a blastocyst of an animal (e.g., a mouse) to form aggregation chimeras. See, e.g., Bradley, 1987. In: TERATOCARCINOMAS AND EMBRYONIC STEM CELLS: A PRACTICAL APPROACH, Robertson, ed. IRL, Oxford, pp. 113-152. A chimeric embryo can then be implanted into a suitable pseudopregnant female foster animal and the embryo brought to term. Progeny harboring the homologously-recombined DNA in their germ cells can
10 be used to breed animals in which all cells of the animal contain the homologously-recombined DNA by germline transmission of the transgene. Methods for constructing homologous recombination vectors and homologous recombinant animals are described further in Bradley, 1991. *Curr. Opin. Biotechnol.* 2: 823-829; PCT International Publication Nos.: WO 90/11354; WO 91/01140; WO 92/0968; and WO 93/04169.

15 In another embodiment, transgenic non-humans animals can be produced that contain selected systems that allow for regulated expression of the transgene. One example of such a system is the cre/loxP recombinase system of bacteriophage P1. For a description of the cre/loxP recombinase system, See, e.g., Lakso, *et al.*, 1992. *Proc. Natl. Acad. Sci. USA* 89: 6232-6236. Another example of a recombinase system is the FLP recombinase system of *Saccharomyces cerevisiae*. See, O'Gorman, *et al.*, 1991. *Science* 251:1351-1355. If a cre/loxP recombinase
20 system is used to regulate expression of the transgene, animals containing transgenes encoding both the Cre recombinase and a selected protein are required. Such animals can be provided through the construction of "double" transgenic animals, e.g., by mating two transgenic animals, one containing a transgene encoding a selected protein and the other containing a transgene
25 encoding a recombinase.

Clones of the non-human transgenic animals described herein can also be produced according to the methods described in Wilmut, *et al.*, 1997. *Nature* 385: 810-813. In brief, a cell (e.g., a somatic cell) from the transgenic animal can be isolated and induced to exit the growth cycle and enter G₀ phase. The quiescent cell can then be fused, e.g., through the use of electrical
30 pulses, to an enucleated oocyte from an animal of the same species from which the quiescent cell is isolated. The reconstructed oocyte is then cultured such that it develops to morula or blastocyte and then transferred to pseudopregnant female foster animal. The offspring borne of

this female foster animal will be a clone of the animal from which the cell (e.g., the somatic cell) is isolated.

Pharmaceutical Compositions

5 The NOVX nucleic acid molecules, NOVX proteins, and anti-NOVX antibodies (also referred to herein as "active compounds") of the invention, and derivatives, fragments, analogs and homologs thereof, can be incorporated into pharmaceutical compositions suitable for administration. Such compositions typically comprise the nucleic acid molecule, protein, or antibody and a pharmaceutically acceptable carrier. As used herein, "pharmaceutically acceptable carrier" is intended to include any and all solvents, dispersion media, coatings, antibacterial and antifungal agents, isotonic and absorption delaying agents, and the like, compatible with pharmaceutical administration. Suitable carriers are described in the most recent edition of Remington's Pharmaceutical Sciences, a standard reference text in the field, which is incorporated herein by reference. Preferred examples of such carriers or diluents include, but are not limited to, water, saline, finger's solutions, dextrose solution, and 5% human serum albumin. Liposomes and non-aqueous vehicles such as fixed oils may also be used. The use of such media and agents for pharmaceutically active substances is well known in the art. Except insofar as any conventional media or agent is incompatible with the active compound, use thereof in the compositions is contemplated. Supplementary active compounds can also be incorporated into the compositions.

15 The antibodies disclosed herein can also be formulated as immunoliposomes. Liposomes containing the antibody are prepared by methods known in the art, such as described in Epstein et al., Proc. Natl. Acad. Sci. USA, 82: 3688 (1985); Hwang et al., Proc. Natl. Acad. Sci. USA, 77: 4030 (1980); and U.S. Pat. Nos. 4,485,045 and 4,544,545. Liposomes with enhanced circulation time are disclosed in U.S. Patent No. 5,013,556.

25 Particularly useful liposomes can be generated by the reverse-phase evaporation method with a lipid composition comprising phosphatidylcholine, cholesterol, and PEG-derivatized phosphatidylethanolamine (PEG-PE). Liposomes are extruded through filters of defined pore size to yield liposomes with the desired diameter. Fab' fragments of the antibody of the present invention can be conjugated to the liposomes as described in Martin et al., J. Biol. Chem., 257: 286-288 (1982) via a disulfide-interchange reaction. A chemotherapeutic agent (such as

Doxorubicin) is optionally contained within the liposome. See Gabrilson et al., J. National Cancer Inst., 81(19): 1484 (1989).

5 A pharmaceutical composition of the invention is formulated to be compatible with its intended route of administration. Examples of routes of administration include parenteral, *e.g.*, intravenous, intradermal, subcutaneous, oral (*e.g.*, inhalation), transdermal (*i.e.*, topical), transmucosal, and rectal administration. Solutions or suspensions used for parenteral, intradermal, or subcutaneous application can include the following components: a sterile diluent such as water for injection, saline solution, fixed oils, polyethylene glycols, glycerine, propylene glycol or other synthetic solvents; antibacterial agents such as benzyl alcohol or methyl parabens; 10 antioxidants such as ascorbic acid or sodium bisulfite; chelating agents such as ethylenediaminetetraacetic acid (EDTA); buffers such as acetates, citrates or phosphates, and agents for the adjustment of tonicity such as sodium chloride or dextrose. The pH can be adjusted with acids or bases, such as hydrochloric acid or sodium hydroxide. The parenteral preparation can be enclosed in ampoules, disposable syringes or multiple dose vials made of glass or plastic. 15

Pharmaceutical compositions suitable for injectable use include sterile aqueous solutions (where water soluble) or dispersions and sterile powders for the extemporaneous preparation of sterile injectable solutions or dispersion. For intravenous administration, suitable carriers include physiological saline, bacteriostatic water, Cremophor EL™ (BASF, Parsippany, N.J.) or 20 phosphate buffered saline (PBS). In all cases, the composition must be sterile and should be fluid to the extent that easy syringeability exists. It must be stable under the conditions of manufacture and storage and must be preserved against the contaminating action of microorganisms such as bacteria and fungi. The carrier can be a solvent or dispersion medium containing, for example, water, ethanol, polyol (for example, glycerol, propylene glycol, and liquid polyethylene glycol, and the like), and suitable mixtures thereof. The proper fluidity can 25 be maintained, for example, by the use of a coating such as lecithin, by the maintenance of the required particle size in the case of dispersion and by the use of surfactants. Prevention of the action of microorganisms can be achieved by various antibacterial and antifungal agents, for example, parabens, chlorobutanol, phenol, ascorbic acid, thimerosal, and the like. In many cases, 30 it will be preferable to include isotonic agents, for example, sugars, polyalcohols such as manitol, sorbitol, sodium chloride in the composition. Prolonged absorption of the injectable

compositions can be thought about by including in the composition an agent which delays absorption, for example, aluminum monostearate and gelatin.

Sterile injectable solutions can be prepared by incorporating the active compound (e.g., a NOVX protein or anti-NOVX antibody) in the required amount in an appropriate solvent with one or a combination of ingredients enumerated above, as required, followed by filtered sterilization. Generally, dispersions are prepared by incorporating the active compound into a sterile vehicle that contains a basic dispersion medium and the required other ingredients from those enumerated above. In the case of sterile powders for the preparation of sterile injectable solutions, methods of preparation are vacuum drying and freeze-drying that yields a powder of the active ingredient plus any additional desired ingredient from a previously sterile-filtered solution thereof.

Oral compositions generally include an inert diluent or an edible carrier. They can be enclosed in gelatin capsules or compressed into tablets. For the purpose of oral therapeutic administration, the active compound can be incorporated with excipients and used in the form of tablets, troches, or capsules. Oral compositions can also be prepared using a fluid carrier for use as a mouthwash, wherein the compound in the fluid carrier is applied orally and swished and expectorated or swallowed. Pharmaceutically compatible binding agents, and/or adjuvant materials can be included as part of the composition. The tablets, pills, capsules, troches and the like can contain any of the following ingredients, or compounds of a similar nature: a binder such as microcrystalline cellulose, gum tragacanth or gelatin; an excipient such as starch or lactose, a disintegrating agent such as alginic acid, Primogel, or corn starch; a lubricant such as magnesium stearate or Sterotes; a glidant such as colloidal silicon dioxide; a sweetening agent such as sucrose or saccharin; or a flavoring agent such as peppermint, methyl salicylate, or orange flavoring.

For administration by inhalation, the compounds are delivered in the form of an aerosol spray from pressured container or dispenser which contains a suitable propellant, e.g., a gas such as carbon dioxide, or a nebulizer.

Systemic administration can also be by transmucosal or transdermal means. For transmucosal or transdermal administration, penetrants appropriate to the barrier to be permeated are used in the formulation. Such penetrants are generally known in the art, and include, for example, for transmucosal administration, detergents, bile salts, and fusidic acid derivatives. Transmucosal administration can be accomplished through the use of nasal sprays or

suppositories. For transdermal administration, the active compounds are formulated into ointments, salves, gels, or creams as generally known in the art.

The compounds can also be prepared in the form of suppositories (*e.g.*, with conventional suppository bases such as cocoa butter and other glycerides) or retention enemas for rectal delivery.

In one embodiment, the active compounds are prepared with carriers that will protect the compound against rapid elimination from the body, such as a controlled release formulation, including implants and microencapsulated delivery systems. Biodegradable, biocompatible polymers can be used, such as ethylene vinyl acetate, polyanhydrides, polyglycolic acid, collagen, polyorthoesters, and polylactic acid. Methods for preparation of such formulations will be apparent to those skilled in the art. The materials can also be obtained commercially from Alza Corporation and Nova Pharmaceuticals, Inc. Liposomal suspensions (including liposomes targeted to infected cells with monoclonal antibodies to viral antigens) can also be used as pharmaceutically acceptable carriers. These can be prepared according to methods known to those skilled in the art, for example, as described in U.S. Patent No. 4,522,811.

It is especially advantageous to formulate oral or parenteral compositions in dosage unit form for ease of administration and uniformity of dosage. Dosage unit form as used herein refers to physically discrete units suited as unitary dosages for the subject to be treated; each unit containing a predetermined quantity of active compound calculated to produce the desired therapeutic effect in association with the required pharmaceutical carrier. The specification for the dosage unit forms of the invention are dictated by and directly dependent on the unique characteristics of the active compound and the particular therapeutic effect to be achieved, and the limitations inherent in the art of compounding such an active compound for the treatment of individuals.

The nucleic acid molecules of the invention can be inserted into vectors and used as gene therapy vectors. Gene therapy vectors can be delivered to a subject by, for example, intravenous injection, local administration (*see, e.g.*, U.S. Patent No. 5,328,470) or by stereotactic injection (*see, e.g.*, Chen, *et al.*, 1994. *Proc. Natl. Acad. Sci. USA* 91: 3054-3057). The pharmaceutical preparation of the gene therapy vector can include the gene therapy vector in an acceptable diluent, or can comprise a slow release matrix in which the gene delivery vehicle is imbedded. Alternatively, where the complete gene delivery vector can be produced intact from recombinant

cells, e.g., retroviral vectors, the pharmaceutical preparation can in one or more cells that produce the gene delivery system.

Antibodies specifically binding a protein of the invention, as well as other molecules identified by the screening assays disclosed herein, can be administered for the treatment of various disorders in the form of pharmaceutical compositions. Principles and considerations involved in preparing such compositions, as well as guidance in the choice of components are provided, for example, in Remington : The Science And Practice Of Pharmacy 19th ed. (Alfonso R. Gennaro, et al., editors) Mack Pub. Co., Easton, Pa. : 1995; Drug Absorption Enhancement : Concepts, Possibilities, Limitations, And Trends, Harwood Academic Publishers, Langhorne, Pa., 1994; and Peptide And Protein Drug Delivery (Advances In Parenteral Sciences, Vol. 4), 1991, M. Dekker, New York. If the antigenic protein is intracellular and whole antibodies are used as inhibitors, internalizing antibodies are preferred. However, liposomes can also be used to deliver the antibody, or an antibody fragment, into cells. Where antibody fragments are used, the smallest inhibitory fragment that specifically binds to the binding domain of the target protein is preferred. For example, based upon the variable-region sequences of an antibody, peptide molecules can be designed that retain the ability to bind the target protein sequence. Such peptides can be synthesized chemically and/or produced by recombinant DNA technology. See, e.g., Marasco *et al.*, 1993 *Proc. Natl. Acad. Sci. USA*, 90: 7889-7893. The formulation herein can also contain more than one active compound as necessary for the particular indication being treated, preferably those with complementary activities that do not adversely affect each other. Alternatively, or in addition, the composition can comprise an agent that enhances its function, such as, for example, a cytotoxic agent, cytokine, chemotherapeutic agent, or growth-inhibitory agent. Such molecules are suitably present in combination in amounts that are effective for the purpose intended. The active ingredients can also be entrapped in microcapsules prepared, for example, by coacervation techniques or by interfacial polymerization, for example, hydroxymethylcellulose or gelatin-microcapsules and poly-(methylmethacrylate) microcapsules, respectively, in colloidal drug delivery systems (for example, liposomes, albumin microspheres, microemulsions, nano-particles, and nanocapsules) or in macroemulsions.

The formulations to be used for *iv vivo* administration must be sterile. This is readily accomplished by filtration through sterile filtration membranes.

Sustained-release preparations can be prepared. Suitable examples of sustained-release preparations include semipermeable matrices of solid hydrophobic polymers containing the

antibody, which matrices are in the form of shaped articles, e.g., films, or microcapsules. Examples of sustained-release matrices include polyesters, hydrogels (for example, poly(2-hydroxyethyl-methacrylate), or poly(vinylalcohol)), polylactides (U.S. Pat. No. 3,773,919), copolymers of L-glutamic acid and ethyl-L-glutamate, non-degradable ethylene-vinyl acetate, 5 degradable lactic acid-glycolic acid copolymers such as the LUPRON DEPOT™ (injectable microspheres composed of lactic acid-glycolic acid copolymer and leuprolide acetate), and poly-D-(-)-3-hydroxybutyric acid. While polymers such as ethylene-vinyl acetate and lactic acid-glycolic acid enable release of molecules for over 100 days, certain hydrogels release proteins for shorter time periods.

- 10 The pharmaceutical compositions can be included in a container, pack, or dispenser together with instructions for administration.

Screening and Detection Methods

- 15 The isolated nucleic acid molecules of the invention can be used to express NOVX protein (e.g., via a recombinant expression vector in a host cell in gene therapy applications), to detect NOVX mRNA (e.g., in a biological sample) or a genetic lesion in a NOVX gene, and to modulate NOVX activity, as described further, below. In addition, the NOVX proteins can be used to screen drugs or compounds that modulate the NOVX protein activity or expression as well as to treat disorders characterized by insufficient or excessive production of NOVX protein 20 or production of NOVX protein forms that have decreased or aberrant activity compared to NOVX wild-type protein. In addition, the anti-NOVX antibodies of the invention can be used to detect and isolate NOVX proteins and modulate NOVX activity. For example, NOVX activity includes growth and differentiation, antibody production, and tumor growth.

- 25 The invention further pertains to novel agents identified by the screening assays described herein and uses thereof for treatments as described, *supra*.

Screening Assays

- 30 The invention provides a method (also referred to herein as a "screening assay") for identifying modulators, *i.e.*, candidate or test compounds or agents (e.g., peptides, peptidomimetics, small molecules or other drugs) that bind to NOVX proteins or have a stimulatory or inhibitory effect on, e.g., NOVX protein expression or NOVX protein activity. The invention also includes compounds identified in the screening assays described herein.

NOVX protein determined. The cell, for example, can be of mammalian origin or a yeast cell. Determining the ability of the test compound to bind to the NOVX protein can be accomplished, for example, by coupling the test compound with a radioisotope or enzymatic label such that binding of the test compound to the NOVX protein or biologically-active portion thereof can be determined by detecting the labeled compound in a complex. For example, test compounds can be labeled with ^{125}I , ^{35}S , ^{14}C , or ^3H , either directly or indirectly, and the radioisotope detected by direct counting of radioemission or by scintillation counting. Alternatively, test compounds can be enzymatically-labeled with, for example, horseradish peroxidase, alkaline phosphatase, or luciferase, and the enzymatic label detected by determination of conversion of an appropriate substrate to product. In one embodiment, the assay comprises contacting a cell which expresses a membrane-bound form of NOVX protein, or a biologically-active portion thereof, on the cell surface with a known compound which binds NOVX to form an assay mixture, contacting the assay mixture with a test compound, and determining the ability of the test compound to interact with a NOVX protein, wherein determining the ability of the test compound to interact with a NOVX protein comprises determining the ability of the test compound to preferentially bind to NOVX protein or a biologically-active portion thereof as compared to the known compound.

In another embodiment, an assay is a cell-based assay comprising contacting a cell expressing a membrane-bound form of NOVX protein, or a biologically-active portion thereof, on the cell surface with a test compound and determining the ability of the test compound to modulate (*e.g.*, stimulate or inhibit) the activity of the NOVX protein or biologically-active portion thereof. Determining the ability of the test compound to modulate the activity of NOVX or a biologically-active portion thereof can be accomplished, for example, by determining the ability of the NOVX protein to bind to or interact with a NOVX target molecule. As used herein, a "target molecule" is a molecule with which a NOVX protein binds or interacts in nature, for example, a molecule on the surface of a cell which expresses a NOVX interacting protein, a molecule on the surface of a second cell, a molecule in the extracellular milieu, a molecule associated with the internal surface of a cell membrane or a cytoplasmic molecule. A NOVX target molecule can be a non-NOVX molecule or a NOVX protein or polypeptide of the invention. In one embodiment, a NOVX target molecule is a component of a signal transduction pathway that facilitates transduction of an extracellular signal (*e.g.* a signal generated by binding of a compound to a membrane-bound NOVX molecule) through the cell membrane and into the

cell. The target, for example, can be a second intercellular protein that has catalytic activity or a protein that facilitates the association of downstream signaling molecules with NOVX.

Determining the ability of the NOVX protein to bind to or interact with a NOVX target molecule can be accomplished by one of the methods described above for determining direct binding. In one embodiment, determining the ability of the NOVX protein to bind to or interact with a NOVX target molecule can be accomplished by determining the activity of the target molecule. For example, the activity of the target molecule can be determined by detecting induction of a cellular second messenger of the target (*i.e.* intracellular Ca^{2+} , diacylglycerol, IP_3 , etc.), detecting catalytic/enzymatic activity of the target an appropriate substrate, detecting the induction of a reporter gene (comprising a NOVX-responsive regulatory element operatively linked to a nucleic acid encoding a detectable marker, *e.g.*, luciferase), or detecting a cellular response, for example, cell survival, cellular differentiation, or cell proliferation.

In yet another embodiment, an assay of the invention is a cell-free assay comprising contacting a NOVX protein or biologically-active portion thereof with a test compound and determining the ability of the test compound to bind to the NOVX protein or biologically-active portion thereof. Binding of the test compound to the NOVX protein can be determined either directly or indirectly as described above. In one such embodiment, the assay comprises contacting the NOVX protein or biologically-active portion thereof with a known compound which binds NOVX to form an assay mixture, contacting the assay mixture with a test compound, and determining the ability of the test compound to interact with a NOVX protein, wherein determining the ability of the test compound to interact with a NOVX protein comprises determining the ability of the test compound to preferentially bind to NOVX or biologically-active portion thereof as compared to the known compound.

In still another embodiment, an assay is a cell-free assay comprising contacting NOVX protein or biologically-active portion thereof with a test compound and determining the ability of the test compound to modulate (*e.g.* stimulate or inhibit) the activity of the NOVX protein or biologically-active portion thereof. Determining the ability of the test compound to modulate the activity of NOVX can be accomplished, for example, by determining the ability of the NOVX protein to bind to a NOVX target molecule by one of the methods described above for determining direct binding. In an alternative embodiment, determining the ability of the test compound to modulate the activity of NOVX protein can be accomplished by determining the ability of the NOVX protein further modulate a NOVX target molecule. For example, the

catalytic/enzymatic activity of the target molecule on an appropriate substrate can be determined as described above.

In yet another embodiment, the cell-free assay comprises contacting the NOVX protein or biologically-active portion thereof with a known compound which binds NOVX protein to form an assay mixture, contacting the assay mixture with a test compound, and determining the ability of the test compound to interact with a NOVX protein, wherein determining the ability of the test compound to interact with a NOVX protein comprises determining the ability of the NOVX protein to preferentially bind to or modulate the activity of a NOVX target molecule.

The cell-free assays of the invention are amenable to use of both the soluble form or the membrane-bound form of NOVX protein. In the case of cell-free assays comprising the membrane-bound form of NOVX protein, it may be desirable to utilize a solubilizing agent such that the membrane-bound form of NOVX protein is maintained in solution. Examples of such solubilizing agents include non-ionic detergents such as n-octylglucoside, n-dodecylglucoside, n-dodecylmaltoside, octanoyl-N-methylglucamide, decanoyl-N-methylglucamide, Triton® X-100, Triton® X-114, Thesit®, Isotridecypoly(ethylene glycol ether)_n, N-dodecyl--N,N-dimethyl-3-ammonio-1-propane sulfonate, 3-(3-cholamidopropyl) dimethylamminiol-1-propane sulfonate (CHAPS), or 3-(3-cholamidopropyl)dimethylamminiol-2-hydroxy-1-propane sulfonate (CHAPSO).

In more than one embodiment of the above assay methods of the invention, it may be desirable to immobilize either NOVX protein or its target molecule to facilitate separation of complexed from uncomplexed forms of one or both of the proteins, as well as to accommodate automation of the assay. Binding of a test compound to NOVX protein, or interaction of NOVX protein with a target molecule in the presence and absence of a candidate compound, can be accomplished in any vessel suitable for containing the reactants. Examples of such vessels include microtiter plates, test tubes, and micro-centrifuge tubes. In one embodiment, a fusion protein can be provided that adds a domain that allows one or both of the proteins to be bound to a matrix. For example, GST-NOVX fusion proteins or GST-target fusion proteins can be adsorbed onto glutathione sepharose beads (Sigma Chemical, St. Louis, MO) or glutathione derivatized microtiter plates, that are then combined with the test compound or the test compound and either the non-adsorbed target protein or NOVX protein, and the mixture is incubated under conditions conducive to complex formation (e.g., at physiological conditions for salt and pH). Following incubation, the beads or microtiter plate wells are washed to remove any

unbound components, the matrix immobilized in the case of beads, complex determined either directly or indirectly, for example, as described, *supra*. Alternatively, the complexes can be dissociated from the matrix, and the level of NOVX protein binding or activity determined using standard techniques.

5 Other techniques for immobilizing proteins on matrices can also be used in the screening assays of the invention. For example, either the NOVX protein or its target molecule can be immobilized utilizing conjugation of biotin and streptavidin. Biotinylated NOVX protein or target molecules can be prepared from biotin-NHS (N-hydroxy-succinimide) using techniques well-known within the art (*e.g.*, biotinylation kit, Pierce Chemicals, Rockford, Ill.), and
10 immobilized in the wells of streptavidin-coated 96 well plates (Pierce Chemical). Alternatively, antibodies reactive with NOVX protein or target molecules, but which do not interfere with binding of the NOVX protein to its target molecule, can be derivatized to the wells of the plate, and unbound target or NOVX protein trapped in the wells by antibody conjugation. Methods for detecting such complexes, in addition to those described above for the GST-immobilized
15 complexes, include immunodetection of complexes using antibodies reactive with the NOVX protein or target molecule, as well as enzyme-linked assays that rely on detecting an enzymatic activity associated with the NOVX protein or target molecule.

In another embodiment, modulators of NOVX protein expression are identified in a method wherein a cell is contacted with a candidate compound and the expression of NOVX
20 mRNA or protein in the cell is determined. The level of expression of NOVX mRNA or protein in the presence of the candidate compound is compared to the level of expression of NOVX mRNA or protein in the absence of the candidate compound. The candidate compound can then be identified as a modulator of NOVX mRNA or protein expression based upon this comparison. For example, when expression of NOVX mRNA or protein is greater (*i.e.*, statistically
25 significantly greater) in the presence of the candidate compound than in its absence, the candidate compound is identified as a stimulator of NOVX mRNA or protein expression. Alternatively, when expression of NOVX mRNA or protein is less (statistically significantly less) in the presence of the candidate compound than in its absence, the candidate compound is identified as an inhibitor of NOVX mRNA or protein expression. The level of NOVX mRNA or
30 protein expression in the cells can be determined by methods described herein for detecting NOVX mRNA or protein.

In yet another aspect of the invention, the NOVX proteins can be used as "bait proteins" in a two-hybrid assay or three hybrid assay (see, e.g., U.S. Patent No. 5,283,317; Zervos, *et al.*, 1993. *Cell* 72: 223-232; Madura, *et al.*, 1993. *J. Biol. Chem.* 268: 12046-12054; Bartel, *et al.*, 1993. *Biotechniques* 14: 920-924; Iwabuchi, *et al.*, 1993. *Oncogene* 8: 1693-1696; and Brent WO 94/10300), to identify other proteins that bind to or interact with NOVX ("NOVX-binding proteins" or "NOVX-bp") and modulate NOVX activity. Such NOVX-binding proteins are also likely to be involved in the propagation of signals by the NOVX proteins as, for example, upstream or downstream elements of the NOVX pathway.

The two-hybrid system is based on the modular nature of most transcription factors, which consist of separable DNA-binding and activation domains. Briefly, the assay utilizes two different DNA constructs. In one construct, the gene that codes for NOVX is fused to a gene encoding the DNA binding domain of a known transcription factor (e.g., GAL-4). In the other construct, a DNA sequence, from a library of DNA sequences, that encodes an unidentified protein ("prey" or "sample") is fused to a gene that codes for the activation domain of the known transcription factor. If the "bait" and the "prey" proteins are able to interact, *in vivo*, forming a NOVX-dependent complex, the DNA-binding and activation domains of the transcription factor are brought into close proximity. This proximity allows transcription of a reporter gene (e.g., LacZ) that is operably linked to a transcriptional regulatory site responsive to the transcription factor. Expression of the reporter gene can be detected and cell colonies containing the functional transcription factor can be isolated and used to obtain the cloned gene that encodes the protein which interacts with NOVX.

The invention further pertains to novel agents identified by the aforementioned screening assays and uses thereof for treatments as described herein.

Detection Assays

Portions or fragments of the cDNA sequences identified herein (and the corresponding complete gene sequences) can be used in numerous ways as polynucleotide reagents. By way of example, and not of limitation, these sequences can be used to: (i) identify an individual from a minute biological sample (tissue typing); and (ii) aid in forensic identification of a biological sample. Some of these applications are described in the subsections, below.

Tissue Typing

The NOVX sequences of the invention can be used to identify individuals from minute biological samples. In this technique, an individual's genomic DNA is digested with one or more
5 restriction enzymes, and probed on a Southern blot to yield unique bands for identification. The sequences of the invention are useful as additional DNA markers for RFLP ("restriction fragment length polymorphisms," described in U.S. Patent No. 5,272,057).

Furthermore, the sequences of the invention can be used to provide an alternative technique that determines the actual base-by-base DNA sequence of selected portions of an
10 individual's genome. Thus, the NOVX sequences described herein can be used to prepare two PCR primers from the 5'- and 3'-termini of the sequences. These primers can then be used to amplify an individual's DNA and subsequently sequence it.

Panels of corresponding DNA sequences from individuals, prepared in this manner, can provide unique individual identifications, as each individual will have a unique set of such DNA
15 sequences due to allelic differences. The sequences of the invention can be used to obtain such identification sequences from individuals and from tissue. The NOVX sequences of the invention uniquely represent portions of the human genome. Allelic variation occurs to some degree in the coding regions of these sequences, and to a greater degree in the noncoding regions. It is estimated that allelic variation between individual humans occurs with a frequency of about
20 once per each 500 bases. Much of the allelic variation is due to single nucleotide polymorphisms (SNPs), which include restriction fragment length polymorphisms (RFLPs).

Each of the sequences described herein can, to some degree, be used as a standard against which DNA from an individual can be compared for identification purposes. Because greater numbers of polymorphisms occur in the noncoding regions, fewer sequences are necessary to
25 differentiate individuals. The noncoding sequences can comfortably provide positive individual identification with a panel of perhaps 10 to 1,000 primers that each yield a noncoding amplified sequence of 100 bases. If predicted coding sequences, such as those in SEQ ID NO: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19 or 21 are used, a more appropriate number of primers for positive individual identification would be 500-2,000.

Predictive Medicine

The invention also pertains to the field of predictive medicine in which diagnostic assays, prognostic assays, pharmacogenomics, and monitoring clinical trials are used for prognostic

- 5 (predictive) purposes to thereby treat an individual prophylactically. Accordingly, one aspect of the invention relates to diagnostic assays for determining NOVX protein and/or nucleic acid expression as well as NOVX activity, in the context of a biological sample (*e.g.*, blood, serum, cells, tissue) to thereby determine whether an individual is afflicted with a disease or disorder, or is at risk of developing a disorder, associated with aberrant NOVX expression or activity.
- 10 Disorders associated with aberrant NOVX expression of activity include, for example, disorders characterized by aberrant cell proliferation, differentiation and migration, *e.g.* cancer, angiogenesis and wound healing, neurological disorders, *e.g.* paraneoplastic neurological disorders, episodic ataxia, autosomal dominant myokymia, stroke, Parkinson's disease and Alzheimer's disease, enamel defects, *e.g.* amelogenesis imperfecta, and inappropriate proteolysis,
- 15 *e.g.* atherosclerosis and abdominal aortic aneurisms.

The invention also provides for prognostic (or predictive) assays for determining whether an individual is at risk of developing a disorder associated with NOVX protein, nucleic acid expression or activity. For example, mutations in a NOVX gene can be assayed in a biological sample. Such assays can be used for prognostic or predictive purpose to thereby prophylactically

20 treat an individual prior to the onset of a disorder characterized by or associated with NOVX protein, nucleic acid expression, or biological activity.

Another aspect of the invention provides methods for determining NOVX protein, nucleic acid expression or activity in an individual to thereby select appropriate therapeutic or prophylactic agents for that individual (referred to herein as "pharmacogenomics").

- 25 Pharmacogenomics allows for the selection of agents (*e.g.*, drugs) for therapeutic or prophylactic treatment of an individual based on the genotype of the individual (*e.g.*, the genotype of the individual examined to determine the ability of the individual to respond to a particular agent.)

Yet another aspect of the invention pertains to monitoring the influence of agents (*e.g.*, drugs, compounds) on the expression or activity of NOVX in clinical trials.

- 30 These and other agents are described in further detail in the following sections.

Diagnostic Assays

An exemplary method for detecting the presence or absence of NOVX in a biological sample involves obtaining a biological sample from a test subject and contacting the biological sample with a compound or an agent capable of detecting NOVX protein or nucleic acid (e.g., mRNA, genomic DNA) that encodes NOVX protein such that the presence of NOVX is detected in the biological sample. An agent for detecting NOVX mRNA or genomic DNA is a labeled nucleic acid probe capable of hybridizing to NOVX mRNA or genomic DNA. The nucleic acid probe can be, for example, a full-length NOVX nucleic acid, such as the nucleic acid of SEQ ID NO: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19 or 21, or a portion thereof, such as an oligonucleotide of at least 15, 30, 50, 100, 250 or 500 nucleotides in length and sufficient to specifically hybridize under stringent conditions to NOVX mRNA or genomic DNA. Other suitable probes for use in the diagnostic assays of the invention are described herein.

One agent for detecting NOVX protein is an antibody capable of binding to NOVX protein, preferably an antibody with a detectable label. Antibodies directed against a protein of the invention may be used in methods known within the art relating to the localization and/or quantitation of the protein (e.g., for use in measuring levels of the protein within appropriate physiological samples, for use in diagnostic methods, for use in imaging the protein, and the like). In a given embodiment, antibodies against the proteins, or derivatives, fragments, analogs or homologs thereof, that contain the antigen binding domain, are utilized as pharmacologically-active compounds.

An antibody specific for a protein of the invention can be used to isolate the protein by standard techniques, such as immunoaffinity chromatography or immunoprecipitation. Such an antibody can facilitate the purification of the natural protein antigen from cells and of recombinantly produced antigen expressed in host cells. Moreover, such an antibody can be used to detect the antigenic protein (e.g., in a cellular lysate or cell supernatant) in order to evaluate the abundance and pattern of expression of the antigenic protein. Antibodies directed against the protein can be used diagnostically to monitor protein levels in tissue as part of a clinical testing procedure, e.g., to, for example, determine the efficacy of a given treatment regimen. Detection can be facilitated by coupling (i.e., physically linking) the antibody to a detectable substance. Examples of detectable substances include various enzymes, prosthetic groups, fluorescent materials, luminescent materials, bioluminescent materials, and radioactive materials. Examples

of suitable enzymes include horseradish peroxidase, alkaline phosphatase, β -galactosidase, or acetylcholinesterase; examples of suitable prosthetic group complexes include streptavidin/biotin and avidin/biotin; examples of suitable fluorescent materials include umbelliferone, fluorescein, fluorescein isothiocyanate, rhodamine, dichlorotriazinylamine fluorescein, dansyl chloride or phycoerythrin; an example of a luminescent material includes luminol; examples of bioluminescent materials include luciferase, luciferin, and aequorin, and examples of suitable radioactive material include ^{125}I , ^{131}I , ^{35}S or ^3H .

Antibodies can be polyclonal, or more preferably, monoclonal. An intact antibody, or a fragment thereof (e.g., Fab or F(ab')₂) can be used. The term "labeled", with regard to the probe or antibody, is intended to encompass direct labeling of the probe or antibody by coupling (i.e., physically linking) a detectable substance to the probe or antibody, as well as indirect labeling of the probe or antibody by reactivity with another reagent that is directly labeled. Examples of indirect labeling include detection of a primary antibody using a fluorescently-labeled secondary antibody and end-labeling of a DNA probe with biotin such that it can be detected with fluorescently-labeled streptavidin. The term "biological sample" is intended to include tissues, cells and biological fluids isolated from a subject, as well as tissues, cells and fluids present within a subject. That is, the detection method of the invention can be used to detect NOVX mRNA, protein, or genomic DNA in a biological sample *in vitro* as well as *in vivo*. For example, *in vitro* techniques for detection of NOVX mRNA include Northern hybridizations and *in situ* hybridizations. *In vitro* techniques for detection of NOVX protein include enzyme linked immunosorbent assays (ELISAs), Western blots, immunoprecipitations, and immunofluorescence. *In vitro* techniques for detection of NOVX genomic DNA include Southern hybridizations. Furthermore, *in vivo* techniques for detection of NOVX protein include introducing into a subject a labeled anti-NOVX antibody. For example, the antibody can be labeled with a radioactive marker whose presence and location in a subject can be detected by standard imaging techniques.

In one embodiment, the biological sample contains protein molecules from the test subject. Alternatively, the biological sample can contain mRNA molecules from the test subject or genomic DNA molecules from the test subject. A preferred biological sample is a peripheral blood leukocyte sample isolated by conventional means from a subject.

In one embodiment, the methods further involve obtaining a control biological sample from a control subject, contacting the control sample with a compound or agent capable of

detecting NOVX protein, mRNA, or genomic DNA, such that the presence of NOVX protein, mRNA or genomic DNA is detected in the biological sample, and comparing the presence of NOVX protein, mRNA or genomic DNA in the control sample with the presence of NOVX protein, mRNA or genomic DNA in the test sample.

5 The invention also encompasses kits for detecting the presence of NOVX in a biological sample. For example, the kit can comprise: a labeled compound or agent capable of detecting NOVX protein or mRNA in a biological sample; means for determining the amount of NOVX in the sample; and means for comparing the amount of NOVX in the sample with a standard. The compound or agent can be packaged in a suitable container. The kit can further comprise
10 instructions for using the kit to detect NOVX protein or nucleic acid.

Prognostic Assays

15 The diagnostic methods described herein can furthermore be utilized to identify subjects having or at risk of developing a disease or disorder associated with aberrant NOVX expression or activity. For example, the assays described herein, such as the preceding diagnostic assays or the following assays, can be utilized to identify a subject having or at risk of developing a disorder associated with NOVX protein, nucleic acid expression or activity. Such disorders include for example, disorders characterized by aberrant cell proliferation, differentiation and migration, *e.g.* cancer, angiogenesis and wound healing, neurological disorders, *e.g.*
20 paraneoplastic neurological disorders, episodic ataxia, autosomal dominant myokymia, stroke, Parkinson's disease and Alzheimer's disease, enamel defects, *e.g.* amelogenesis imperfecta, and inappropriate proteolysis, *e.g.* atherosclerosis and abdominal aortic aneurisms.

25 Alternatively, the prognostic assays can be utilized to identify a subject having or at risk for developing a disease or disorder. Thus, the invention provides a method for identifying a disease or disorder associated with aberrant NOVX expression or activity in which a test sample is obtained from a subject and NOVX protein or nucleic acid (*e.g.*, mRNA, genomic DNA) is detected, wherein the presence of NOVX protein or nucleic acid is diagnostic for a subject having or at risk of developing a disease or disorder associated with aberrant NOVX expression or activity. As used herein, a "test sample" refers to a biological sample obtained from a subject
30 of interest. For example, a test sample can be a biological fluid (*e.g.*, serum), cell sample, or tissue.

Furthermore, the prognostic assays described herein can be used to determine whether a subject can be administered an agent (*e.g.*, an agonist, antagonist, peptidomimetic, protein, peptide, nucleic acid, small molecule, or other drug candidate) to treat a disease or disorder associated with aberrant NOVX expression or activity. For example, such methods can be used to determine whether a subject can be effectively treated with an agent for a disorder. Thus, the invention provides methods for determining whether a subject can be effectively treated with an agent for a disorder associated with aberrant NOVX expression or activity in which a test sample is obtained and NOVX protein or nucleic acid is detected (*e.g.*, wherein the presence of NOVX protein or nucleic acid is diagnostic for a subject that can be administered the agent to treat a disorder associated with aberrant NOVX expression or activity).

The methods of the invention can also be used to detect genetic lesions in a NOVX gene, thereby determining if a subject with the lesioned gene is at risk for a disorder characterized by aberrant cell proliferation and/or differentiation. In various embodiments, the methods include detecting, in a sample of cells from the subject, the presence or absence of a genetic lesion characterized by at least one of an alteration affecting the integrity of a gene encoding a NOVX-protein, or the misexpression of the NOVX gene. For example, such genetic lesions can be detected by ascertaining the existence of at least one of: (i) a deletion of one or more nucleotides from a NOVX gene; (ii) an addition of one or more nucleotides to a NOVX gene; (iii) a substitution of one or more nucleotides of a NOVX gene, (iv) a chromosomal rearrangement of a NOVX gene; (v) an alteration in the level of a messenger RNA transcript of a NOVX gene, (vi) aberrant modification of a NOVX gene, such as of the methylation pattern of the genomic DNA, (vii) the presence of a non-wild-type splicing pattern of a messenger RNA transcript of a NOVX gene, (viii) a non-wild-type level of a NOVX protein, (ix) allelic loss of a NOVX gene, and (x) inappropriate post-translational modification of a NOVX protein. As described herein, there are a large number of assay techniques known in the art which can be used for detecting lesions in a NOVX gene. A preferred biological sample is a peripheral blood leukocyte sample isolated by conventional means from a subject. However, any biological sample containing nucleated cells may be used, including, for example, buccal mucosal cells.

In certain embodiments, detection of the lesion involves the use of a probe/primer in a polymerase chain reaction (PCR) (*see, e.g.*, U.S. Patent Nos. 4,683,195 and 4,683,202), such as anchor PCR or RACE PCR, or, alternatively, in a ligation chain reaction (LCR) (*see, e.g.*, Landegran, *et al.*, 1988. *Science* 241: 1077-1080; and Nakazawa, *et al.*, 1994. *Proc. Natl. Acad.*

Sci. USA 91: 360-364), the latter of which can be particularly useful for detecting point mutations in the NOVX-gene (see, Abravaya, *et al.*, 1995. *Nucl. Acids Res.* 23: 675-682). This method can include the steps of collecting a sample of cells from a patient, isolating nucleic acid (*e.g.*, genomic, mRNA or both) from the cells of the sample, contacting the nucleic acid sample with one or more primers that specifically hybridize to a NOVX gene under conditions such that hybridization and amplification of the NOVX gene (if present) occurs, and detecting the presence or absence of an amplification product, or detecting the size of the amplification product and comparing the length to a control sample. It is anticipated that PCR and/or LCR may be desirable to use as a preliminary amplification step in conjunction with any of the techniques used for detecting mutations described herein.

Alternative amplification methods include: self sustained sequence replication (see, Guatelli, *et al.*, 1990. *Proc. Natl. Acad. Sci. USA* 87: 1874-1878), transcriptional amplification system (see, Kwoh, *et al.*, 1989. *Proc. Natl. Acad. Sci. USA* 86: 1173-1177); Q β Replicase (see, Lizardi, *et al.*, 1988. *BioTechnology* 6: 1197), or any other nucleic acid amplification method, followed by the detection of the amplified molecules using techniques well known to those of skill in the art. These detection schemes are especially useful for the detection of nucleic acid molecules if such molecules are present in very low numbers.

In an alternative embodiment, mutations in a NOVX gene from a sample cell can be identified by alterations in restriction enzyme cleavage patterns. For example, sample and control DNA is isolated, amplified (optionally), digested with one or more restriction endonucleases, and fragment length sizes are determined by gel electrophoresis and compared. Differences in fragment length sizes between sample and control DNA indicates mutations in the sample DNA. Moreover, the use of sequence specific ribozymes (see, *e.g.*, U.S. Patent No. 5,493,531) can be used to score for the presence of specific mutations by development or loss of a ribozyme cleavage site.

In other embodiments, genetic mutations in NOVX can be identified by hybridizing a sample and control nucleic acids, *e.g.*, DNA or RNA, to high-density arrays containing hundreds or thousands of oligonucleotide probes. See, *e.g.*, Cronin, *et al.*, 1996. *Human Mutation* 7: 244-255; Kozal, *et al.*, 1996. *Nat. Med.* 2: 753-759. For example, genetic mutations in NOVX can be identified in two dimensional arrays containing light-generated DNA probes as described in Cronin, *et al.*, *supra*. Briefly, a first hybridization array of probes can be used to scan through long stretches of DNA in a sample and control to identify base changes between the sequences

by making linear arrays of sequential overlapping probes. This step allows the identification of point mutations. This is followed by a second hybridization array that allows the characterization of specific mutations by using smaller, specialized probe arrays complementary to all variants or mutations detected. Each mutation array is composed of parallel probe sets, one complementary to the wild-type gene and the other complementary to the mutant gene.

In yet another embodiment, any of a variety of sequencing reactions known in the art can be used to directly sequence the NOVX gene and detect mutations by comparing the sequence of the sample NOVX with the corresponding wild-type (control) sequence. Examples of

sequencing reactions include those based on techniques developed by Maxim and Gilbert, 1977. *Proc. Natl. Acad. Sci. USA* 74: 560 or Sanger, 1977. *Proc. Natl. Acad. Sci. USA* 74: 5463. It is also contemplated that any of a variety of automated sequencing procedures can be utilized when performing the diagnostic assays (see, e.g., Naeve, et al., 1995. *Biotechniques* 19: 448), including sequencing by mass spectrometry (see, e.g., PCT International Publication No. WO 94/16101; Cohen, et al., 1996. *Adv. Chromatography* 36: 127-162; and Griffin, et al., 1993. *Appl. Biochem. Biotechnol.* 38: 147-159).

Other methods for detecting mutations in the NOVX gene include methods in which protection from cleavage agents is used to detect mismatched bases in RNA/RNA or RNA/DNA heteroduplexes. See, e.g., Myers, et al., 1985. *Science* 230: 1242. In general, the art technique of "mismatch cleavage" starts by providing heteroduplexes of formed by hybridizing (labeled) RNA or DNA containing the wild-type NOVX sequence with potentially mutant RNA or DNA obtained from a tissue sample. The double-stranded duplexes are treated with an agent that cleaves single-stranded regions of the duplex such as which will exist due to basepair mismatches between the control and sample strands. For instance, RNA/DNA duplexes can be treated with RNase and DNA/DNA hybrids treated with S₁ nuclease to enzymatically digesting the mismatched regions. In other embodiments, either DNA/DNA or RNA/DNA duplexes can be treated with hydroxylamine or osmium tetroxide and with piperidine in order to digest mismatched regions. After digestion of the mismatched regions, the resulting material is then separated by size on denaturing polyacrylamide gels to determine the site of mutation. See, e.g., Cotton, et al., 1988. *Proc. Natl. Acad. Sci. USA* 85: 4397; Saleeba, et al., 1992. *Methods Enzymol.* 217: 286-295. In an embodiment, the control DNA or RNA can be labeled for detection.

In still another embodiment, the mismatch cleavage reaction employs one or more proteins that recognize mismatched base pairs in double-stranded DNA (so called "DNA mismatch repair" enzymes) in defined systems for detecting and mapping point mutations in NOVX cDNAs obtained from samples of cells. For example, the mutY enzyme of *E. coli* cleaves A at G/A mismatches and the thymidine DNA glycosylase from HeLa cells cleaves T at G/T mismatches. See, e.g., Hsu, *et al.*, 1994. *Carcinogenesis* 15: 1657-1662. According to an exemplary embodiment, a probe based on a NOVX sequence, e.g., a wild-type NOVX sequence, is hybridized to a cDNA or other DNA product from a test cell(s). The duplex is treated with a DNA mismatch repair enzyme, and the cleavage products, if any, can be detected from electrophoresis protocols or the like. See, e.g., U.S. Patent No. 5,459,039.

In other embodiments, alterations in electrophoretic mobility will be used to identify mutations in NOVX genes. For example, single strand conformation polymorphism (SSCP) may be used to detect differences in electrophoretic mobility between mutant and wild type nucleic acids. See, e.g., Orita, *et al.*, 1989. *Proc. Natl. Acad. Sci. USA*: 86: 2766; Cotton, 1993. *Mutat. Res.* 285: 125-144; Hayashi, 1992. *Genet. Anal. Tech. Appl.* 9: 73-79. Single-stranded DNA fragments of sample and control NOVX nucleic acids will be denatured and allowed to renature. The secondary structure of single-stranded nucleic acids varies according to sequence, the resulting alteration in electrophoretic mobility enables the detection of even a single base change. The DNA fragments may be labeled or detected with labeled probes. The sensitivity of the assay may be enhanced by using RNA (rather than DNA), in which the secondary structure is more sensitive to a change in sequence. In one embodiment, the subject method utilizes heteroduplex analysis to separate double stranded heteroduplex molecules on the basis of changes in electrophoretic mobility. See, e.g., Keen, *et al.*, 1991. *Trends Genet.* 7: 5.

In yet another embodiment, the movement of mutant or wild-type fragments in polyacrylamide gels containing a gradient of denaturant is assayed using denaturing gradient gel electrophoresis (DGGE). See, e.g., Myers, *et al.*, 1985. *Nature* 313: 495. When DGGE is used as the method of analysis, DNA will be modified to insure that it does not completely denature, for example by adding a GC clamp of approximately 40 bp of high-melting GC-rich DNA by PCR. In a further embodiment, a temperature gradient is used in place of a denaturing gradient to identify differences in the mobility of control and sample DNA. See, e.g., Rosenbaum and Reissner, 1987. *Biophys. Chem.* 265: 12753.

Examples of other techniques for detecting point mutations include, but are not limited to, selective oligonucleotide hybridization, selective amplification, or selective primer extension. For example, oligonucleotide primers may be prepared in which the known mutation is placed centrally and then hybridized to target DNA under conditions that permit hybridization only if a perfect match is found. See, e.g., Saiki, et al., 1986. *Nature* 324: 163; Saiki, et al., 1989. *Proc. Natl. Acad. Sci. USA* 86: 6230. Such allele specific oligonucleotides are hybridized to PCR amplified target DNA or a number of different mutations when the oligonucleotides are attached to the hybridizing membrane and hybridized with labeled target DNA.

Alternatively, allele specific amplification technology that depends on selective PCR amplification may be used in conjunction with the instant invention. Oligonucleotides used as primers for specific amplification may carry the mutation of interest in the center of the molecule (so that amplification depends on differential hybridization; see, e.g., Gibbs, et al., 1989. *Nucl. Acids Res.* 17: 2437-2448) or at the extreme 3'-terminus of one primer where, under appropriate conditions, mismatch can prevent, or reduce polymerase extension (see, e.g., Prossner, 1993. *Tibtech.* 11: 238). In addition it may be desirable to introduce a novel restriction site in the region of the mutation to create cleavage-based detection. See, e.g., Gasparini, et al., 1992. *Mol. Cell Probes* 6: 1. It is anticipated that in certain embodiments amplification may also be performed using *Taq* ligase for amplification. See, e.g., Barany, 1991. *Proc. Natl. Acad. Sci. USA* 88: 189. In such cases, ligation will occur only if there is a perfect match at the 3'-terminus of the 5' sequence, making it possible to detect the presence of a known mutation at a specific site by looking for the presence or absence of amplification.

The methods described herein may be performed, for example, by utilizing pre-packaged diagnostic kits comprising at least one probe nucleic acid or antibody reagent described herein, which may be conveniently used, e.g., in clinical settings to diagnose patients exhibiting symptoms or family history of a disease or illness involving a NOVX gene.

Furthermore, any cell type or tissue, preferably peripheral blood leukocytes, in which NOVX is expressed may be utilized in the prognostic assays described herein. However, any biological sample containing nucleated cells may be used, including, for example, buccal mucosal cells.

Pharmacogenomics

Agents, or modulators that have a stimulatory or inhibitory effect on NOVX activity (e.g., NOVX gene expression), as identified by a screening assay described herein can be administered

to individuals to treat (prophylactically or therapeutically) disorders characterized by aberrant cell proliferation, differentiation and migration, e.g. cancer, angiogenesis and wound healing, neurological disorders, e.g. paraneoplastic neurological disorders, episodic ataxia, autosomal dominant myokymia, stroke, Parkinson's disease and Alzheimer's disease, enamel defects, e.g. amelogenesis imperfecta, and inappropriate proteolysis, e.g. atherosclerosis and abdominal aortic aneurisms. In conjunction with such treatment, the pharmacogenomics (i.e., the study of the relationship between an individual's genotype and that individual's response to a foreign compound or drug) of the individual may be considered. Differences in metabolism of therapeutics can lead to severe toxicity or therapeutic failure by altering the relation between dose and blood concentration of the pharmacologically active drug. Thus, the pharmacogenomics of the individual permits the selection of effective agents (e.g., drugs) for prophylactic or therapeutic treatments based on a consideration of the individual's genotype. Such pharmacogenomics can further be used to determine appropriate dosages and therapeutic regimens. Accordingly, the activity of NOVX protein, expression of NOVX nucleic acid, or mutation content of NOVX genes in an individual can be determined to thereby select appropriate agent(s) for therapeutic or prophylactic treatment of the individual.

Pharmacogenomics deals with clinically significant hereditary variations in the response to drugs due to altered drug disposition and abnormal action in affected persons. See e.g., Eichelbaum, 1996. *Clin. Exp. Pharmacol. Physiol.*, 23: 983-985; Linder, 1997. *Clin. Chem.*, 43: 254-266. In general, two types of pharmacogenetic conditions can be differentiated. Genetic conditions transmitted as a single factor altering the way drugs act on the body (altered drug action) or genetic conditions transmitted as single factors altering the way the body acts on drugs (altered drug metabolism). These pharmacogenetic conditions can occur either as rare defects or as polymorphisms. For example, glucose-6-phosphate dehydrogenase (G6PD) deficiency is a common inherited enzymopathy in which the main clinical complication is hemolysis after ingestion of oxidant drugs (anti-malarials, sulfonamides, analgesics, nitrofurans) and consumption of fava beans.

As an illustrative embodiment, the activity of drug metabolizing enzymes is a major determinant of both the intensity and duration of drug action. The discovery of genetic polymorphisms of drug metabolizing enzymes (e.g., N-acetyltransferase 2 (NAT 2) and cytochrome P450 enzymes CYP2D6 and CYP2C19) has provided an explanation as to why some patients do not obtain the expected drug effects or show exaggerated drug response and serious

toxicity after taking the standard and safe dose of a drug. These polymorphisms are expressed in two phenotypes in the population, the extensive metabolizer (EM) and poor metabolizer (PM). The prevalence of PM is different among different populations. For example, the gene coding for CYP2D6 is highly polymorphic and several mutations have been identified in PM, which all lead to the absence of functional CYP2D6. Poor metabolizers of CYP2D6 and CYP2C19 quite frequently experience exaggerated drug response and side effects when they receive standard doses. If a metabolite is the active therapeutic moiety, PM show no therapeutic response, as demonstrated for the analgesic effect of codeine mediated by its CYP2D6-formed metabolite morphine. At the other extreme are the so called ultra-rapid metabolizers who do not respond to standard doses. Recently, the molecular basis of ultra-rapid metabolism has been identified to be due to CYP2D6 gene amplification.

Thus, the activity of NOVX protein, expression of NOVX nucleic acid, or mutation content of NOVX genes in an individual can be determined to thereby select appropriate agent(s) for therapeutic or prophylactic treatment of the individual. In addition, pharmacogenetic studies can be used to apply genotyping of polymorphic alleles encoding drug-metabolizing enzymes to the identification of an individual's drug responsiveness phenotype. This knowledge, when applied to dosing or drug selection, can avoid adverse reactions or therapeutic failure and thus enhance therapeutic or prophylactic efficiency when treating a subject with a NOVX modulator, such as a modulator identified by one of the exemplary screening assays described herein.

Monitoring of Effects During Clinical Trials

Monitoring the influence of agents (*e.g.*, drugs, compounds) on the expression or activity of NOVX (*e.g.*, the ability to modulate aberrant cell proliferation) can be applied not only in basic drug screening, but also in clinical trials. For example, the effectiveness of an agent determined by a screening assay as described herein to increase NOVX gene expression, protein levels, or upregulate NOVX activity, can be monitored in clinical trials of subjects exhibiting decreased NOVX gene expression, protein levels, or downregulated NOVX activity. Alternatively, the effectiveness of an agent determined by a screening assay to decrease NOVX gene expression, protein levels, or downregulate NOVX activity, can be monitored in clinical trails of subjects exhibiting increased NOVX gene expression, protein levels, or upregulated NOVX activity. In such clinical trials, the expression or activity of NOVX and, preferably, other

genes that have been implicated in, for example, a cellular proliferation or immune disorder can be used as a "read out" or markers of the immune responsiveness of a particular cell.

By way of example, and not of limitation, genes, including NOVX, that are modulated in cells by treatment with an agent (*e.g.*, compound, drug or small molecule) that modulates NOVX activity (*e.g.*, identified in a screening assay as described herein) can be identified. Thus, to study the effect of agents on cellular proliferation disorders, for example, in a clinical trial, cells can be isolated and RNA prepared and analyzed for the levels of expression of NOVX and other genes implicated in the disorder. The levels of gene expression (*i.e.*, a gene expression pattern) can be quantified by Northern blot analysis or RT-PCR, as described herein, or alternatively by measuring the amount of protein produced, by one of the methods as described herein, or by measuring the levels of activity of NOVX or other genes. In this manner, the gene expression pattern can serve as a marker, indicative of the physiological response of the cells to the agent. Accordingly, this response state may be determined before, and at various points during, treatment of the individual with the agent.

In one embodiment, the invention provides a method for monitoring the effectiveness of treatment of a subject with an agent (*e.g.*, an agonist, antagonist, protein, peptide, peptidomimetic, nucleic acid, small molecule, or other drug candidate identified by the screening assays described herein) comprising the steps of (i) obtaining a pre-administration sample from a subject prior to administration of the agent; (ii) detecting the level of expression of a NOVX protein, mRNA, or genomic DNA in the preadministration sample; (iii) obtaining one or more post-administration samples from the subject; (iv) detecting the level of expression or activity of the NOVX protein, mRNA, or genomic DNA in the post-administration samples; (v) comparing the level of expression or activity of the NOVX protein, mRNA, or genomic DNA in the pre-administration sample with the NOVX protein, mRNA, or genomic DNA in the post administration sample or samples; and (vi) altering the administration of the agent to the subject accordingly. For example, increased administration of the agent may be desirable to increase the expression or activity of NOVX to higher levels than detected, *i.e.*, to increase the effectiveness of the agent. Alternatively, decreased administration of the agent may be desirable to decrease expression or activity of NOVX to lower levels than detected, *i.e.*, to decrease the effectiveness of the agent.

Methods of Treatment

The invention provides for both prophylactic and therapeutic methods of treating a subject at risk of (or susceptible to) a disorder or having a disorder associated with aberrant NOVX expression or activity. Disorders associated with aberrant NOVX expression include, for example, disorders characterized by aberrant cell proliferation, differentiation and migration, *e.g.* cancer, angiogenesis and wound healing, neurological disorders, *e.g.* paraneoplastic neurological disorders, episodic ataxia, autosomal dominant myokymia, stroke, Parkinson's disease and Alzheimer's disease, enamel defects, *e.g.* amelogenesis imperfecta, and inappropriate proteolysis, *e.g.* atherosclerosis and abdominal aortic aneurisms. These methods of treatment will be discussed more fully, below.

Disease and Disorders

Diseases and disorders that are characterized by increased (relative to a subject not suffering from the disease or disorder) levels or biological activity may be treated with Therapeutics that antagonize (*i.e.*, reduce or inhibit) activity. Therapeutics that antagonize activity may be administered in a therapeutic or prophylactic manner. Therapeutics that may be utilized include, but are not limited to: (i) an aforementioned peptide, or analogs, derivatives, fragments or homologs thereof; (ii) antibodies to an aforementioned peptide; (iii) nucleic acids encoding an aforementioned peptide; (iv) administration of antisense nucleic acid and nucleic acids that are "dysfunctional" (*i.e.*, due to a heterologous insertion within the coding sequences of coding sequences to an aforementioned peptide) that are utilized to "knockout" endogenous function of an aforementioned peptide by homologous recombination (*see, e.g.*, Capecchi, 1989. *Science* 244: 1288-1292); or (v) modulators (*i.e.*, inhibitors, agonists and antagonists, including additional peptide mimetic of the invention or antibodies specific to a peptide of the invention) that alter the interaction between an aforementioned peptide and its binding partner.

Diseases and disorders that are characterized by decreased (relative to a subject not suffering from the disease or disorder) levels or biological activity may be treated with Therapeutics that increase (*i.e.*, are agonists to) activity. Therapeutics that upregulate activity may be administered in a therapeutic or prophylactic manner. Therapeutics that may be utilized include, but are not limited to, an aforementioned peptide, or analogs, derivatives, fragments or homologs thereof; or an agonist that increases bioavailability.

Increased or decreased levels can be readily detected by quantifying peptide and/or RNA, by obtaining a patient tissue sample (e.g., from biopsy tissue) and assaying it *in vitro* for RNA or peptide levels, structure and/or activity of the expressed peptides (or mRNAs of an aforementioned peptide). Methods that are well-known within the art include, but are not limited to, immunoassays (e.g., by Western blot analysis, immunoprecipitation followed by sodium dodecyl sulfate (SDS) polyacrylamide gel electrophoresis, immunocytochemistry, etc.) and/or hybridization assays to detect expression of mRNAs (e.g., Northern assays, dot blots, *in situ* hybridization, and the like).

10 Prophylactic Methods

In one aspect, the invention provides a method for preventing, in a subject, a disease or condition associated with an aberrant NOVX expression or activity, by administering to the subject an agent that modulates NOVX expression or at least one NOVX activity. Subjects at risk for a disease that is caused or contributed to by aberrant NOVX expression or activity can be identified by, for example, any or a combination of diagnostic or prognostic assays as described herein. Administration of a prophylactic agent can occur prior to the manifestation of symptoms characteristic of the NOVX aberrancy, such that a disease or disorder is prevented or, alternatively, delayed in its progression. Depending upon the type of NOVX aberrancy, for example, a NOVX agonist or NOVX antagonist agent can be used for treating the subject. The appropriate agent can be determined based on screening assays described herein. The prophylactic methods of the invention are further discussed in the following subsections.

Therapeutic Methods

Another aspect of the invention pertains to methods of modulating NOVX expression or activity for therapeutic purposes. The modulatory method of the invention involves contacting a cell with an agent that modulates one or more of the activities of NOVX protein activity associated with the cell. An agent that modulates NOVX protein activity can be an agent as described herein, such as a nucleic acid or a protein, a naturally-occurring cognate ligand of a NOVX protein, a peptide, a NOVX peptidomimetic, or other small molecule. In one embodiment, the agent stimulates one or more NOVX protein activity. Examples of such stimulatory agents include active NOVX protein and a nucleic acid molecule encoding NOVX that has been introduced into the cell. In another embodiment, the agent inhibits one or more

NOVX protein activity. Examples of such inhibitory agents include antisense NOVX nucleic acid molecules and anti-NOVX antibodies. These modulatory methods can be performed *in vitro* (e.g., by culturing the cell with the agent) or, alternatively, *in vivo* (e.g., by administering the agent to a subject). As such, the invention provides methods of treating an individual afflicted with a disease or disorder characterized by aberrant expression or activity of a NOVX protein or nucleic acid molecule. In one embodiment, the method involves administering an agent (e.g., an agent identified by a screening assay described herein), or combination of agents that modulates (e.g., up-regulates or down-regulates) NOVX expression or activity. In another embodiment, the method involves administering a NOVX protein or nucleic acid molecule as therapy to compensate for reduced or aberrant NOVX expression or activity.

Stimulation of NOVX activity is desirable in situations in which NOVX is abnormally downregulated and/or in which increased NOVX activity is likely to have a beneficial effect. One example of such a situation is where a subject has a disorder characterized by aberrant cell proliferation and/or differentiation (e.g., cancer or immune associated). Another example of such a situation is where the subject has an immunodeficiency disease (e.g., AIDS).

Antibodies of the invention, including polyclonal, monoclonal, humanized and fully human antibodies, may be used as therapeutic agents. Such agents will generally be employed to treat or prevent a disease or pathology in a subject. An antibody preparation, preferably one having high specificity and high affinity for its target antigen, is administered to the subject and will generally have an effect due to its binding with the target. Such an effect may be one of two kinds, depending on the specific nature of the interaction between the given antibody molecule and the target antigen in question. In the first instance, administration of the antibody may abrogate or inhibit the binding of the target with an endogenous ligand to which it naturally binds. In this case, the antibody binds to the target and masks a binding site of the naturally occurring ligand, wherein the ligand serves as an effector molecule. Thus the receptor mediates a signal transduction pathway for which ligand is responsible.

Alternatively, the effect may be one in which the antibody elicits a physiological result by virtue of binding to an effector binding site on the target molecule. In this case the target, a receptor having an endogenous ligand which may be absent or defective in the disease or pathology, binds the antibody as a surrogate effector ligand, initiating a receptor-based signal transduction event by the receptor.

A therapeutic effective amount of an antibody of the invention relates generally to the amount needed to achieve a therapeutic objective. As noted above, this may be a binding interaction between the antibody and its target antigen that, in certain cases, interferes with the functioning of the target, and in other cases, promotes a physiological response. The amount required to be administered will furthermore depend on the binding affinity of the antibody for its specific antigen, and will also depend on the rate at which an administered antibody is depleted from the free volume other subject to which it is administered. Common ranges for therapeutically effective dosing of an antibody or antibody fragment of the invention may be, by way of nonlimiting example, from about 0.1 mg/kg body weight to about 50 mg/kg body weight. Common dosing frequencies may range, for example, from twice daily to once a week.

Determination of the Biological Effect of the Therapeutic

In various embodiments of the invention, suitable *in vitro* or *in vivo* assays are performed to determine the effect of a specific Therapeutic and whether its administration is indicated for treatment of the affected tissue.

In various specific embodiments, *in vitro* assays may be performed with representative cells of the type(s) involved in the patient's disorder, to determine if a given Therapeutic exerts the desired effect upon the cell type(s). Compounds for use in therapy may be tested in suitable animal model systems including, but not limited to rats, mice, chicken, cows, monkeys, rabbits, and the like, prior to testing in human subjects. Similarly, for *in vivo* testing, any of the animal model system known in the art may be used prior to administration to human subjects.

The invention will be further described in the following examples, which do not limit the scope of the invention described in the claims.

EXAMPLES

Example 1.: Method of Identifying the Nucleic Acids of the Present Invention.

Novel nucleic acid sequences were identified by TblastN using CuraGen Corporation's sequence file run against the Genomic Daily Files made available by GenBank. The nucleic acids were further predicted by the program GenScan™, including selection of exons. These were further modified by means of similarities using BLAST searches. The sequences were then

manually corrected for apparent inconsistencies, thereby obtaining the sequences encoding the full-length protein.

Example 2.: Identification of NOV3 (CG51785-06)

5 The sequence of Acc. No. CG51785-06 was derived by laboratory cloning of cDNA fragments, by *in silico* prediction of the sequence. cDNA fragments covering either the full length of the DNA sequence, or part of the sequence, or both, were cloned. In silico prediction was based on sequences available in Curagen's proprietary sequence databases or in the public human sequence databases, and provided either the full length DNA sequence, or some portion
10 thereof.

The laboratory cloning was performed using one or more of the methods summarized below:

SeqCalling™ Technology: cDNA was derived from various human samples representing multiple tissue types, normal and diseased states, physiological states, and developmental states
15 from different donors. Samples were obtained as whole tissue, primary cells or tissue cultured primary cells or cell lines. Cells and cell lines may have been treated with biological or chemical agents that regulate gene expression, for example, growth factors, chemokines or steroids. The cDNA thus derived was then sequenced using CuraGen's proprietary SeqCalling technology. Sequence traces were evaluated manually and edited for corrections if appropriate. cDNA
20 sequences from all samples were assembled together, sometimes including public human sequences, using bioinformatic programs to produce a consensus sequence for each assembly. Each assembly is included in CuraGen Corporation's database. Sequences were included as components for assembly when the extent of identity with another component was at least 95% over 50 bp. Each assembly represents a gene or portion thereof and includes information on
25 variants, such as splice forms single nucleotide polymorphisms (SNPs), insertions, deletions and other sequence variations.

Exon Linking: The cDNA coding for the CG51785-06 sequence was cloned by the polymerase chain reaction (PCR) using the primers: 5'-TCTCCCACAGGCCAGGAC-3' (SEQ ID NO.: 59) and 5'-CGCATGGTTTTGGGATTG-3' (SEQ ID NO.: 60). Primers were designed
30 based on in silico predictions of the full length or some portion (one or more exons) of the cDNA/protein sequence of the invention. These primers were used to amplify a cDNA from a

pool containing expressed human sequences derived from the following tissues: adrenal gland, bone marrow, brain - amygdala, brain - cerebellum, brain - hippocampus, brain - substantia nigra, brain - thalamus, brain -whole, fetal brain, fetal kidney, fetal liver, fetal lung, heart, kidney, lymphoma - Raji, mammary gland, pancreas, pituitary gland, placenta, prostate, salivary gland, skeletal muscle, small intestine, spinal cord, spleen, stomach, testis, thyroid, trachea and uterus.

Multiple clones were sequenced and these fragments were assembled together, sometimes including public human sequences, using bioinformatic programs to produce a consensus sequence for each assembly. Each assembly is included in CuraGen Corporation's database. Sequences were included as components for assembly when the extent of identity with another component was at least 95% over 50 bp. Each assembly represents a gene or portion thereof and includes information on variants, such as splice forms single nucleotide polymorphisms (SNPs), insertions, deletions and other sequence variations.

Physical clone: The PCR product derived by exon linking, covering the entire open reading frame, was cloned into the pCR2.1 vector from Invitrogen to provide clone 27824582_0_105.698496.F5.

Variant sequences are also included in this application. A variant sequence can include a single nucleotide polymorphism (SNP). A SNP can, in some instances, be referred to as a "cSNP" to denote that the nucleotide sequence containing the SNP originates as a cDNA. A SNP can arise in several ways. For example, a SNP may be due to a substitution of one nucleotide for another at the polymorphic site. Such a substitution can be either a transition or a transversion. A SNP can also arise from a deletion of a nucleotide or an insertion of a nucleotide, relative to a reference allele. In this case, the polymorphic site is a site at which one allele bears a gap with respect to a particular nucleotide in another allele. SNPs occurring within genes may result in an alteration of the amino acid encoded by the gene at the position of the SNP. Intragenic SNPs may also be silent, when a codon including a SNP encodes the same amino acid as a result of the redundancy of the genetic code. SNPs occurring outside the region of a gene, or in an intron within a gene, do not result in changes in any amino acid sequence of a protein but may result in altered regulation of the expression pattern. Examples include alteration in temporal expression, physiological response regulation, cell type expression regulation, intensity of expression, and stability of transcribed message.

The DNA sequence and protein sequence for a novel Collagen-like gene were obtained by exon linking and are reported here as NOV3 (CG51785-06).

Example 3. Molecular cloning of NOV11 (AL096677_A)

5

A mature form of a NOV11 protein predicted for clone AL096677_A (SEQ ID NO: 22), namely, the region from amino acid residue 20 to residue 148 was targeted for cloning. The PCR primers shown below were prepared.

10 AL096677_A Mat-F: GGATCCGCCAAGCTGGGTCACCTTCCAAAGGTGG (SEQ
ID NO: 61), and
AL096677_A REV: CTCGAGTCTGAGGTTTCTGCCACATGCTCGG (SEQ
ID NO: 62).

15 A PCR reaction was set up using 5 ng human testis cDNA template. The reaction mixtures contained 1 microM of each of the AL096677_A Mat-F and AL096677_A REV primers, 5 micromoles dNTP (Clontech Laboratories, Palo Alto CA) and 1 microliter of 50xAdvantage-HF 2 polymerase (Clontech Laboratories, Palo Alto CA) in 50 microliter reaction volume. The following reaction conditions were used:

- 20 a) 96°C 3 minutes
b) 96°C 30 seconds denaturation
c) 70°C 30 seconds, primer annealing. This temperature was gradually decreased by 1°C/cycle
d) 72°C 3 minutes extension.
25 Repeat steps b-d 10 times
e) 96°C 30 seconds denaturation
f) 60°C 30 seconds annealing
g) 72°C 3 minutes extension
Repeat steps e-g 35 times
30 h) 72°C 5 minutes final extension

An amplified product having the expected size of approximately 450 bp was detected by agarose gel electrophoresis. The fragment was purified from agarose gel and ligated to pCR2.1 vector (Invitrogen, Carlsbad, CA) following the manufacturer's recommendation. The clone is called
35 pCR2.1-AL096655_A-S602-9B. The cloned insert was sequenced, using the following sequence-specific primers:

AL096677_A-S1: GTGGAGTATATAGTCACTGTG (SEQ ID NO: 63) and
AL096677_A-S2: CACAGTGACTATATACTCGAG (SEQ ID NO: 64).

40

The nucleotide sequence obtained for clone pCR2.1-AL096655_A-S602-9B is shown in Table 34.

45 TABLE 34.

GCCAAGCTGGGTCTTCCAAAGGTGGGAGGGCTTCCAGAGAAGCTCATGAGCAA
 GAAGAACATGAATTCAACACTCAACTTCTTCATTCAATCCTACAACAATGCCAGCAA
 CGACACCTACTTATATCGAGTCCAGAGGCTAATTCGAAGTCAGATGCAGCTGACGAC
 GGGAGTGGAGTATATAGTCACTGTGAAGATTGGCCGGACCAAATGCAAGAGGAATG
 5 ACACGAGCAATTCTTCCTGCCCCCTGCAAAGCAAGAAGCTGAGAAAGAGTTTAATTT
 GCGAGTCTTTGATATACACCATGCCCTGGATAAACTATTTCCAGCTCTGGAACAATT
 CCTGTCTGGAGGCCGAGCATGTGGGCAGAAACCTCAGA (SEQ ID NO: 65)

The corresponding amino acid sequence predicted for clone pCR2.1-AL096655_A-S602-9B is
 shown in Table 35.

Table 35.

AKLGHFQRWEGFQQKLMSKKNMNSTLNFFIQSYNNASNDTYLYRVQRLIRSQMQLTTG
 15 VEYIVTVKIGRTKCKRNDTSNSSCPLQSKKLKSLICESLIYTMPWINYFQLWNNSCLEA
 EHVGRNLR (SEQ ID NO. 66)

There are 5 nucleotide changes (in bold underlining) introducing 4 amino acid changes (in bold
 underlining) in the clone compared to the sequence of the corresponding portion of clone
 AL096655_A. These are characterized in Tables 36 and 37.

TABLE 36. Comparison of nucleotide sequences of clone AL09677_A (top row) and clone
 AL09677_A-S602-9B (bottom row).

67	GCCAAGCTGGGTCACTTCCAAAGGTGGGAGGGCTTCCAGCAGAAGCTCAT	116
1	GCCAAGCTGGGTCACTTCCAAAGGTGGGAGGGCTTCCAGCAGAAGCTCAT	50
117	GAGCAAGAAGAACATGAATTCAACACTCAACTTCTTCATTCAATCCTACA	166
51	GAGCAAGAAGAACATGAATTCAACACTCAACTTCTTCATTCAATCCTACA	100
167	ACAATGCCAGCAACGACACCTACTTATATCGAGTCCAGAGGCTAATTCGA	216
101	ACAATGCCAGCAACGACACCTACTTATATCGAGTCCAGAGGCTAATTCGA	150
217	AGTCAGATGCAGCTGACGACGGGAGTGGAGTATATAGTCACTGTGAAGAT	266
151	AGTCAGATGCAGCTGACGACGGGAGTGGAGTATATAGTCACTGTGAAGAT	200
267	TGGCTGGACCAAATGCAAGAGGAATGACACGAGCAATTCTTCCTGCCCCC	316
201	TGGCCGGACCAAATGCAAGAGGAATGACACGAGCAATTCTTCCTGCCCCC	250
317	TGCAAA CA AGAAGCTGAGAAAGAGTTTAATTTGCGAGTCTTTAATATAC	366
251	TGCAAA GA AGAAGCTGAGAAAGAGTTTAATTTGCGAGTCTTTGATATAC	300
367	ACCATGCCCTGGT T AAACTATTTCCAGCTCTGGAACAATTCCTGTCTGGA	416
301	ACCATGCCCTGG A TAAACTATTTCCAGCTCTGGAACAATTCCTGTCTGGA	350

5

10

|||||

15

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100 101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118 119 120 121 122 123 124 125 126 127 128 129 130 131 132 133 134 135 136 137 138 139 140 141 142 143 144 145 146 147 148 149 150 151 152 153 154 155 156 157 158 159 160 161 162 163 164 165 166 167 168 169 170 171 172 173 174 175 176 177 178 179 180 181 182 183 184 185 186 187 188 189 190 191 192 193 194 195 196 197 198 199 200 201 202 203 204 205 206 207 208 209 210 211 212 213 214 215 216 217 218 219 220 221 222 223 224 225 226 227 228 229 230 231 232 233 234 235 236 237 238 239 240 241 242 243 244 245 246 247 248 249 250 251 252 253 254 255 256 257 258 259 260 261 262 263 264 265 266 267 268 269 270 271 272 273 274 275 276 277 278 279 280 281 282 283 284 285 286 287 288 289 290 291 292 293 294 295 296 297 298 299 300 301 302 303 304 305 306 307 308 309 310 311 312 313 314 315 316 317 318 319 320 321 322 323 324 325 326 327 328 329 330 331 332 333 334 335 336 337 338 339 340 341 342 343 344 345 346 347 348 349 350 351 352 353 354 355 356 357 358 359 360 361 362 363 364 365 366 367 368 369 370 371 372 373 374 375 376 377 378 379 380 381 382 383 384 385 386 387 388 389 390 391 392 393 394 395 396 397 398 399 400 401 402 403 404 405 406 407 408 409 410 411 412 413 414 415 416 417 418 419 420 421 422 423 424 425 426 427 428 429 430 431 432 433 434 435 436 437 438 439 440 441 442 443 444 445 446 447 448 449 450 451 452 453 454 455 456 457 458 459 460 461 462 463 464 465 466 467 468 469 470 471 472 473 474 475 476 477 478 479 480 481 482 483 484 485 486 487 488 489 490 491 492 493 494 495 496 497 498 499 500 501 502 503 504 505 506 507 508 509 510 511 512 513 514 515 516 517 518 519 520 521 522 523 524 525 526 527 528 529 530 531 532 533 534 535 536 537 538 539 540 541 542 543 544 545 546 547 548 549 550 551 552 553 554 555 556 557 558 559 560 561 562 563 564 565 566 567 568 569 570 571 572 573 574 575 576 577 578 579 580 581 582 583 584 585 586 587 588 589 590 591 592 593 594 595 596 597 598 599 600 601 602 603 604 605 606 607 608 609 610 611 612 613 614 615 616 617 618 619 620 621 622 623 624 625 626 627 628 629 630 631 632 633 634 635 636 637 638 639 640 641 642 643 644 645 646 647 648 649 650 651 652 653 654 655 656 657 658 659 660 661 662 663 664 665 666 667 668 669 670 671 672 673 674 675 676 677 678 679 680 681 682 683 684 685 686 687 688 689 690 691 692 693 694 695 696 697 698 699 700 701 702 703 704 705 706 707 708 709 710 711 712 713 714 715 716 717 718 719 720 721 722 723 724 725 726 727 728 729 730 731 732 733 734 735 736 737 738 739 740 741 742 743 744 745 746 747 748 749 750 751 752 753 754 755 756 757 758 759 760 761 762 763 764 765 766 767 768 769 770 771 772 773 774 775 776 777 778 779 780 781 782 783 784 785 786 787 788 789 790 791 792 793 794 795 796 797 798 799 800 801 802 803 804 805 806 807 808 809 810 811 812 813 814 815 816 817 818 819 820 821 822 823 824 825 826 827 828 829 830 831 832 833 834 835 836 837 838 839 840 841 842 843 844 845 846 847 848 849 850 851 852 853 854 855 856 857 858 859 860 861 862 863 864 865 866 867 868 869 870 871 872 873 874 875 876 877 878 879 880 881 882 883 884 885 886 887 888 889 890 891 892 893 894 895 896 897 898 899 900 901 902 903 904 905 906 907 908 909 910 911 912 913 914 915 916 917 918 919 920 921 922 923 924 925 926 927 928 929 930 931 932 933 934 935 936 937 938 939 940 941 942 943 944 945 946 947 948 949 950 951 952 953 954 955 956 957 958 959 960 961 962 963 964 965 966 967 968 969 970 971 972 973 974 975 976 977 978 979 980 981 982 983 984 985 986 987 988 989 990 991 992 993 994 995 996 997 998 999 1000 1001 1002 1003 1004 1005 1006 1007 1008 1009 1010 1011 1012 1013 1014 1015 1016 1017 1018 1019 1020 1021 1022 1023 1024 1025 1026 1027 1028 1029 1030 1031 1032 1033 1034 1035 1036 1037 1038 1039 1040 1

20

11-11-61 : 11-11-61

25

118

[illegible]